

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: November 2, 2005, 22:10:58 ; Search time 876.5 Seconds
(without alignments)
6590.438 Million cell updates/sec

Title: US-10-664-859-1

Perfect score: 12037

Sequence: 1 acgagtgctctcttattatc.....aaaaaaaaaaaaaa 6909

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1865214 segs, 418043040 residues

Total number of hits satisfying chosen parameters: 3730428

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10664859/runat_02112005_171806_3577/app_query.fasta_1.7047
-DB=Published Applications_AA -QFMT=fastan -SUFFIX=rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=biosum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10664859_@CGN_1_1_982_@runat_02112005_171806_3577
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US10E_PUBCOMB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
19: /cgn2_6/ptodata/2/pubpaa/US11A_PUBCOMB.pep:*
20: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep:*
21: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
22: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|---------------------------------------|
| 1 | 7040 | 58.5 | 1429 | 20 | US-11-097-143-3129 Sequence 3129, Ap |
| 2 | 570 | 4.7 | 112 | 9 | US-09-915-543-10 Sequence 10, Appl |
| 3 | 570 | 4.7 | 112 | 14 | US-10-322-579-10 Sequence 10, Appl |
| 4 | 570 | 4.7 | 112 | 15 | US-10-664-859-10 Sequence 2178, Ap |
| 5 | 321.5 | 2.7 | 1435 | 15 | US-10-276-774-2178 Sequence 18, Appl |
| 6 | 314 | 2.6 | 1394 | 15 | US-10-381-247B-18 Sequence 1764, Ap |
| 7 | 314 | 2.6 | 1394 | 16 | US-10-723-860-1764 Sequence 5163, Ap |
| 8 | 314 | 2.6 | 1394 | 18 | US-10-756-149-5163 Sequence 15, Appl |
| 9 | 312 | 2.6 | 1426 | 9 | US-09-915-543-15 Sequence 15, Appl |
| 10 | 312 | 2.6 | 1426 | 14 | US-10-322-579-15 Sequence 15, Appl |
| 11 | 312 | 2.6 | 1426 | 15 | US-10-664-859-15 Sequence 1958, Ap |
| 12 | 312 | 2.6 | 1426 | 16 | US-10-408-765A-1958 Sequence 17, Appl |
| 13 | 296.5 | 2.5 | 1450 | 15 | US-10-381-247B-17 Sequence 2, Appli |
| 14 | 296.5 | 2.5 | 1494 | 15 | US-10-381-247B-2 Sequence 17856, A |
| 15 | 269.5 | 2.2 | 1594 | 20 | US-11-097-143-17856 Sequence 27555, A |
| 16 | 258.5 | 2.1 | 1366 | 20 | US-11-097-143-27555 Sequence 7050, Ap |
| 17 | 258.5 | 2.1 | 2151 | 20 | US-11-097-143-7050 Sequence 8729, Ap |
| 18 | 256.5 | 2.1 | 5322 | 17 | US-10-732-923-8729 Sequence 101, App |
| 19 | 247.5 | 2.1 | 2285 | 18 | US-10-773-446-101 Sequence 38793, A |
| 20 | 242.5 | 2.0 | 1161 | 20 | US-11-097-143-38793 Sequence 7035, Ap |
| 21 | 241.5 | 2.0 | 1778 | 20 | US-11-097-143-7035 Sequence 11742, A |
| 22 | 241 | 2.0 | 2280 | 20 | US-11-097-143-8001 Sequence 8001, Ap |
| 23 | 240 | 2.0 | 2112 | 20 | US-11-097-143-8001 Sequence 1964, Ap |
| 24 | 233.5 | 1.9 | 1243 | 16 | US-10-408-765A-1964 Sequence 2, Appli |
| 25 | 233.5 | 1.9 | 1441 | 15 | US-10-355-218-2 Sequence 16689, A |
| 26 | 233.5 | 1.9 | 1441 | 18 | US-10-503-050A-2 Sequence 18448, A |
| 27 | 233 | 1.9 | 2703 | 20 | US-11-097-143-16689 Sequence 90, Appl |
| 28 | 233 | 1.9 | 3190 | 17 | US-10-732-923-18448 Sequence 38103, A |
| 29 | 230.5 | 1.9 | 3275 | 18 | US-10-840-060-90 Sequence 18447, A |
| 30 | 230.5 | 1.9 | 3275 | 20 | US-11-097-143-38103 Sequence 3, Appli |
| 31 | 230 | 1.9 | 3276 | 17 | US-10-732-923-18447 Sequence 8, Appli |
| 32 | 229.5 | 1.9 | 2005 | 9 | US-09-735-367B-3 Sequence 47506, A |
| 33 | 228.5 | 1.9 | 2063 | 9 | US-09-735-367B-2 Sequence 347, App |
| 34 | 228.5 | 1.9 | 2063 | 18 | US-10-973-858-8 Sequence 350, App |
| 35 | 228.5 | 1.9 | 2063 | 15 | US-10-450-763-47506 Sequence 360, App |
| 36 | 228 | 1.9 | 1424 | 15 | US-10-418-027-1 Sequence 355, App |
| 37 | 228 | 1.9 | 1424 | 16 | US-10-473-127-347 Sequence 6418, Ap |
| 38 | 228 | 1.9 | 1424 | 16 | US-10-473-127-350 Sequence 2217, Ap |
| 39 | 228 | 1.9 | 1424 | 16 | US-10-473-127-360 Sequence 236, App |
| 40 | 223.5 | 1.9 | 1522 | 16 | US-10-473-127-355 Sequence 35, Appl |
| 41 | 222.5 | 1.8 | 1441 | 14 | US-10-170-682-3 Sequence 3, Appli |
| 42 | 222.5 | 1.8 | 1459 | 15 | US-10-369-493-6418 Sequence 2, Appli |
| 43 | 222.5 | 1.8 | 1966 | 16 | US-10-408-765A-2217 Sequence 236, App |
| 44 | 222.5 | 1.8 | 2440 | 15 | US-10-341-434-236 Sequence 35, Appl |
| 45 | 222.5 | 1.8 | 2440 | 18 | US-10-885-977-35 |

ALIGNMENTS

RESULT 1
US-11-097-143-3129
; Sequence 3129, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; FILE REFERENCE: DROSOPHILA-GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/457,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769

; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3129
; LENGTH: 1429
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-11-097-143-3129

Alignment Scores:
Pred. No.: 0 Length: 1429
Score: 7040.00 Matches: 1425
Percent Similarity: 71.09% Conservative: 1
Best Local Similarity: 71.04% Mismatches: 3
Query Match: 58.49% Indels: 577
DB: 20 Gaps: 6

US-10-664-859-1 (1-6909) x US-11-097-143-3129 (1-1429)

OY 453 ATGCTCTGCACACATATGCCCCGAGTCCAAACCCACACAGCCGCAACCAACTCCGAT 512
Db 1 MetLeuSerThrThrMetProArgSerProThrGlnGlnInProGlnProAsnSerAsp 20
OY 513 GCCTCCTCAACAAGTGCATCTGGATCAATCCTGGAGCAGCGGATCGGAATGGGACTCG 572
Db 21 AlaSerSerThrSerAlaSerGlySerAsnProGlyAlaAlaIleGlyAsnGlyAspSer 40
OY 573 GCGGCGAGCAGAGTTCTCCGAAGACCTTAATAGCGAACCTTTCTACTTTGTGCGCG 632
Db 41 AlaAlaSerArgSerSerProLysThrLeuAsnSerGluProPheSerThrLeu-Ser-- 59
OY 633 GGTAAGACTGTATGTATTCTTCTTGTGCGGAATTATTAACAACCTTTCTGTTCAGAG 692
Db 60 -----ProAs 61
OY 693 TCAATAATAAATTGACGCCAGAGAAGGACTGAGAAAAGCGACTATCAACTAGTGATTA 752
Db 61 pGlnIleLysLeuThrProGluGluGlyThrGluLysSerGlyLeuSerThrSerAspLys 81
OY 753 AGCTGCCACTGAGAGAGCCCCAGCGAGTGAATAATCTGCCCGAGGAGCAAACTATGCT 812
Db 81 BAAlaAlaThrGlyGlyAlaProGlySerGlyAsnAsnLeuProGluGlyGlnThrMetLe 101
OY 813 AAGCAGAACTCTACGACGACCAATCAACTCGTCCCTAGTCGCTTCTCCACAAAACCTCCAG 872
Db 101 uArgGlnAsnSerThrSerThrIleAsnSerCysLeuValAlaSerProGlnAsnSerSe 121
OY 873 TGAACACTCGAATAGCAGCAATGTGTCTGCTACAGTGGGCTTACTCAGATGTTAGATTG 932
Db 121 rGluHisSerAsnSerSerAsnValSerAlaThrValGlyLeuThrGlnMetValAspCys 141
OY 933 TGACGAGCAATCGAAGAAAACAATGTAGTGTGAAGAGCAGAGGAAGCTGTAAAGACTGC 992
Db 141 sAspGluGlnSerLysLysAsnLysCysSerValLysAspGluGluAla----- 157
OY 993 CCTACAATGGTTAAAAATTTTAAATGTATTGGCGTTCACTTTGTTAATCATTTAATT 1052
Db 157 ----- 157
OY 1053 GTTTTTTTTTTGTATACTTACAATTTTAGTTTAAACTGTAAACTTGACTAAAACTCG 1112
Db 157 ----- 157
OY 1113 CGAAGCTCGGATCAAAAACAGACATTTTCTTGGAACCGTAATTAGCTCATAAAAATATTTA 1172

Db 157 ----- 157
OY 1173 ATTCACTTGATGGAATGCATATCATAGATGTACTCAAAACATCTCAAGAAAGCCTCAAA 1232
Db 157 ----- 157
OY 1233 TTGATCAACTAATTAGTTTGAGAAAAATTGCTGTACTTTTAAAGATATATTAAATTAA 1292
Db 157 ----- 157
OY 1293 AAATTTGCTGAGTGAATGATATAATAGTCACAATAATTTTAACTGCTAAAGCA 1352
Db 157 ----- 157
OY 1353 TTTTGAATAGCCGCTGCTACGAGATGCTACTAGACCGCGGTGTAAGCTAATTTTATT 1412
Db 157 ----- 157
OY 1413 AAAAGCTGTCTAATATTCCATTAACCATTAATGTCCCATTTGAGAAATAAGTTCTAATTA 1472
Db 158 -----GluIleSerSerAsnLys 163
OY 1473 AGCAAAAGTCAAGCAGCTGTGTCGGGCTGCGAAACAGGTTCTACATCCAGTTTGACTGT 1532
Db 163 sAlaLysGlyGlnAlaIleGlyGlyCysGluThrGlySerThrSerSerLeuThrIva 183
OY 1533 CAAGGAGAACCCACCGATGTCTTAGCGAGTTAGTAATATGAAAAAAGAGAAAGAGA 1592
Db 183 lLysGluGluProThrAspValIleuGlySerLeuValAsnMetLysGluGluArgGly 203
OY 1593 AAATCATTCGCCAACGATGTCCCTGTGTTGTTGTTCAATTGTAATGCACAGAGCAA 1652
Db 203 uAsnHisSerProThrMetSerProValGlyPheGlySerIleGlyAsnAlaGlnAspAs 223
OY 1653 CTCGCTACACCGGTAAGTTTAAAGATCCATATAAAGCAATAACAAGATTAAATGT 1712
Db 223 nSerAlaThr----- 226
OY 1713 CAGTTACCAATTTATTGTATGATCAAGAAGAACTACTATAGCGATATCTCCTGCTTTAA 1772
Db 226 ----- 226
OY 1773 TTTTATTTTAATTAGGAATAACGAATATTCTAATTTGTAATAATAATTGATTAATTAA 1832
Db 226 ----- 226
OY 1833 CTAGAAATTAAAAACCTTTGAATTAGGACATACCCCTTCCAAAATCAGTAATCATTTGGG 1892
Db 226 ----- 226
OY 1893 AACGAGGTGTGTCCCGAAGGAGACTACTATAAAACCTTTTGAGCTATCTGATACTGCA 1952
Db 226 ----- 226
OY 1953 CGCTACTAAAATGATTAGTTTAGGAAAATGGGTGAATTTTGTAGGAAGTTTTCATTTT 2012
Db 226 ----- 226
OY 2013 AGAAGAAATGTGATTATTTTATTTAAACCCCTTCAAGCGGAATACTATTGTTCTACGATA 2072
Db 226 ----- 226
OY 2073 TTTTGGAAAAACAATGTGTTAGTTGAAAAGTGCCCTATATAAACAGAAATTCACGGTTTCA 2132
Db 226 ----- 226
OY 2133 AATACTAACGAGTTTGTGATTTAATTTTGATTAAATGAGAAATTATCACACTTCAGTTA 2192
Db 226 ----- 226
OY 2193 AAATGTTTAATTCGATTAAGGTGCGACAATCACAGCAGATTTCCATTTTTCGCGTATAT 2252
Db 226 ----- 226

| | | | | | | | | | | | |
|----|------|---------|--------------|-------------------------|--------------------------------|------------------|-----------------|--------|--------|------------|------|
| QY | 2253 | ATAGAA | GTCCGCTTCA | CACTCTTCTTGCGCGGCTTCA | CCACTACGTGAGTTCCGCGCGCA | 2312 | | | | | |
| Db | 226 | ----- | ----- | ----- | ----- | 226 | | | | | |
| QY | 2313 | GTGATT | TATATAGATG | ATTTACGAGTTATTTTATG | TGTATTTTAATAATAT | 2372 | | | | | |
| Db | 226 | ----- | ----- | ----- | ----- | 226 | | | | | |
| QY | 2373 | CTTATT | TATTCATTTTAC | ATAGTTAAATTTGAAGAATTTCA | AACGACAGTACCGGAA | 2432 | | | | | |
| Db | 227 | ----- | ----- | Pro--Val | LysIleGluArgIleSerAsnAspSerThr | ThrGlu | 240 | | | | |
| QY | 2433 | AAAAA | AGATCGTCC | TTGCAATG | AATATGACGAATG | AGCATGGAAGGCTGC | AATCAG | 2492 | | | |
| Db | 241 | Lys | LysGlySer | SerLeu | ThrMetAsnAsnAspGlu | MetSerMetGlu | GlyCysAsnGln | 260 | | | |
| QY | 2493 | TTGAAT | CCCCGATT | TTTATCAATGAATCTT | TAATAATTCCTGCAATTTG | GAGCATATTAGTA | 2552 | | | | |
| Db | 261 | Leu | AsnProAsp | PheIleAsnGlu | SerLeuAsnAsnProAla | IleSerSerIle | LeuVal | 280 | | | |
| QY | 2553 | AGCGAG | TAGGACCA | ATAACCCGAATCGAGTTG | AGCGGGACGGGAAATTTT | TGACT | 2612 | | | | |
| Db | 281 | Ser | GlyValGly | ProIlePro | GlyIleGlyValGlyAla | GlyThrGly | AsnLeuLeuThr | 300 | | | |
| QY | 2613 | GCCAAC | GCCAATGGA | ATCTCCTCGG | TAGCAGTAATGTTTG | GATTACATGCA | CAGCAA | 2672 | | | |
| Db | 301 | Ala | AsnAlaAsn | GlyIleSer | SerGlySerSerAsn | CysLeuAspTyr | MetGlnGlnGln | 320 | | | |
| QY | 2673 | AATCACA | TATTCGTGTTT | TCAACTCAGCTGG | CCAAAGGGCCGGAATCAG | TTTAAAGC | 2732 | | | | |
| Db | 321 | Asn | IleIlePhe | ValPheSerThr | GlnLeuAlaAsnLysGlyAla | GluSerValLeu | Ser | 340 | | | |
| QY | 2733 | GGTCAAT | TTTCAACTAT | TATGCGTATCAGTCA | CTGCACTGCTACAA | AAAGCTTCCTG | 2792 | | | | |
| Db | 341 | Gly | GlnPheGln | ThrIleIleAla | TyrHisCysThrGln | ProAlaThrLys | SerPheLeu | 360 | | | |
| QY | 2793 | GAAGACT | TTTTTATGA | AAACCCTTTAA | GATTAACAAGTTAC | AGCGGCACAAT | TCGTC | 2852 | | | |
| Db | 361 | Glu | AspPhePhe | MetLysAsnPro | LeuLysIleAsnLys | LeuGlnArgHis | AsnSerVal | 380 | | | |
| QY | 2853 | GGTATG | CCCATGAT | GAGCATGGGGCAG | GTGCACTAATCCTCA | TATCCTGAGCCAA | 2912 | | | | |
| Db | 381 | Gly | MetPro | TyrIleGly | MetGlyGlnValGly | LeuThrProPro | AsnProValAla | Lys | 400 | | |
| QY | 2913 | ATAACA | CAACA | CAGCACCACAT | ACAAGACCGTAG | CGCTATTGAAACCC | CAATTCAA | 2972 | | | |
| Db | 401 | Ile | ThrGlnGln | GlnProHisThr | LysThrValGly | LeuLeuLysPro | GlnPheAsnGln | 420 | | | |
| QY | 2973 | CATGA | AAAA | CAGCAACGTAG | TACTGTAA | CGCGCTAGCAACTCTT | TGTGCACAGTCT | 3032 | | | |
| Db | 421 | His | GluAsn | SerLysArgSer | ThrValSerAla | ProSerAsnSerPhe | ValAspGlnSer | 440 | | | |
| QY | 3033 | GATCCT | ATGGGCA | ACGAAC | TGAATTGATGTG | CTGGGAAGCGGATCCT | CAAAACCA | 3092 | | | |
| Db | 441 | Asp | Pro | MetGlyAsn | GluThrGlyLeu | MetCysTyrGlyGly | LysSerSerAsnThr | Ser | 460 | | |
| QY | 3093 | AGGTCT | GGACAAA | ACTCAGAA | ATCATGTAGACAG | TATCATCATCCAG | CGAGTCA | CAG | 3152 | | |
| Db | 461 | Arg | SerGlyGln | AsnSerArgAsn | HisValAspSerIle | SerThrSerSer | GluSerGln | 480 | | | |
| QY | 3153 | GCAATA | AAAGAT | ACTGGAAG | CAGCTGCGTTG | ATTGGGACAGGTC | ACAAAGGA | CGAT | 3212 | | |
| Db | 481 | Ala | IleLysIle | LeuGluAlaAla | GlyValAsp | LeuGlyGlnValThr | LysGlySerAsp | 500 | | | |
| QY | 3213 | CCTG | CGCTGA | CACTGA | AAACA | CATTTGTATCACTG | CAAGAGTTAAG | TTCCAG | CGAA | 3272 | |
| Db | 501 | Pro | GlyLeu | ThrThrGlu | AsnAsnIleVal | IleSerLeu | GlnGlyValIle | ValPro | AspGlu | 520 | |
| QY | 3273 | AACCT | TACACCA | ACA | CGGGCA | CATCGGGA | GAACA | AGTTGG | CAAAA | ATAAAAAATG | 3332 |
| Db | 521 | Asn | LeuThrPro | GlnGlnArg | GlnHisArg | GluGluGlnLeu | AlaLysIleLys | LysMet | 540 | | |

| | | | |
|----|------|---|------|
| QY | 3333 | AATCAATTCTTTTCTCCTGAAATGAGAATTCAGTAGGAGCTAATGTAGCTCACAGATA | 3392 |
| | | | |
| Db | 541 | AsnGlnPheLeuPheProGlnAsnGlnAsnSerValGlyAlaAsnValSerSerGlnIle | 560 |
| QY | 3393 | ACAAAAATTCAGAGATTAAATGATGGGATGTCGGGTGGCCGAGCGGATCTATTATA | 3452 |
| | | | |
| Db | 561 | ThrLysIleProGlnAspLeuMetMetGlyMetSerGlyGlyGlySerIleIle | 580 |
| QY | 3453 | AATCCGAGATGCGACAACTGCATATGCCAGGTAAAGCCAAATCGAGCTCTTATCGCGC | 3512 |
| | | | |
| Db | 581 | AsnProThiMetArgGlnLeuHisMetProGlnAsnAlaLysSerGluLeuLeuSerAla | 600 |
| QY | 3513 | ACAAGTTCAGGACTTTCGGAAGATGTAATGCATCCAGGGATGTTATATCAGATATGGGT | 3572 |
| | | | |
| Db | 601 | ThrSerSerGlyLeuSerGluAspValMetHisProGlnAspValIleSerAspMetGly | 620 |
| QY | 3573 | GCCGTAATAGGATGTAATAATCAAAAAACCAGTGTCCAATGTGGATCTGAGTAGGT | 3632 |
| | | | |
| Db | 621 | AlaValIleGlyCysAsnAsnAsnGlnLysThrSerValGlnCysGlySerGlyValGly | 640 |
| QY | 3633 | GTTGTCACTGGAAACAACACTGCAGCTGGAGTAAATGTCAAATATGCATTTGCTCAAGCTCCGGC | 3692 |
| | | | |
| Db | 641 | ValValThrGlyThrThrAlaAlaGlyValAlaAsnValAsnMetHisCysSerSerSerGly | 660 |
| QY | 3693 | GCCCCGAGTCGCAATATGATGGGAAGCTCTACGGATATGCTAGCCCTGTTGGCAACACA | 3752 |
| | | | |
| Db | 661 | AlaProAsnGlnAsnMetMetGlySerSerThrAspMetLeuAlaSerPheGlyAsnThr | 680 |
| QY | 3753 | AGCTGCAACGTTCATCGGAACGCGCCAGATATGTCTTAAGGAAGTTTAAATCAAGATAGC | 3812 |
| | | | |
| Db | 681 | SerCysAsnValIleGlyThrAlaProAspMetSerLysGluValLeuAsnGlnAspSer | 700 |
| QY | 3813 | CGAACCCATTCACATCAAGGGGAGTTGCTCAAAATGAGTGTGCGAAGATTCACATCAA | 3872 |
| | | | |
| Db | 701 | ArgThrHisSerHisGlnGlyGlyValAlaGlnMetGluTrpSerLysIleGlnHisGln | 720 |
| QY | 3873 | TTTTTCGAGAAGACGCTCAAGGGGGGCAAGCCGACAGCAAGTCACTGGAAGTGTAGTACCA | 3932 |
| | | | |
| Db | 721 | PhePheGlnGluArgLeuLysGlyGlyLysProArgGlnValThrGlyThrValValPro | 740 |
| QY | 3933 | CAACAGCAAAACCCCTTCTGGAATCTGTGGAAACTCGTTAAACAACGAGTGGACCCCTG | 3992 |
| | | | |
| Db | 741 | GlnGlnGlnThrProSerGlySerGlyGlyAsnSerLeuAsnAsnGlnValArgProLeu | 760 |
| QY | 3993 | CAAGTTCACCTCCTCCTTACACTCCATCCAGAGATCTGCGTCAGTACCAATAGCCACT | 4052 |
| | | | |
| Db | 761 | GlnGlyProProProProIleHisSerIleGlnArgSerAlaSerValProIleAlaThr | 780 |
| QY | 4053 | CAATCGCCCAATCCCTCGAGTCCAAACAATCTATCTCTCCCGTCACCCGGGACAACCGCA | 4112 |
| | | | |
| Db | 781 | GlnSerProAsnProSerSerProAsnAsnLeuSerLeuProSerProArgThrThrAla | 800 |
| QY | 4113 | GCAGTCATGGGATTGGCCGACCACTCTCCTAGCATGAGTGAACACAGATCATTTACTGGA | 4172 |
| | | | |
| Db | 801 | AlaValMetGlyLeuProThiAsnSerProSerMetAspGlyThrGlySerLeuSerGly | 820 |
| QY | 4173 | TCTGTTCCGCAAGCTAATACTTCAGCGGTTCAAGGACGACACAACAACAGTGTCTCAGCA | 4232 |
| | | | |
| Db | 821 | SerValProGlnAlaAsnThrSerThrValGlnAlaGlyThrThrValLeuSerAla | 840 |
| QY | 4233 | AACAAGAACTGTTTTCAAGGACACACCCCATCGCCGTCAAATCAAAATCGTAGTAGAAAT | 4292 |
| | | | |
| Db | 841 | AsnLysAsnCysPheGlnAlaAspThrProSerProSerAsnGlnAsnArgSerArgAsn | 860 |
| QY | 4293 | ACCGGATCGTCAAGCGTCTTACGCATATACTTAAGACACAACCCCAAGTACCCCTTATCT | 4352 |
| | | | |
| Db | 861 | ThrGlySerSerSerValLeuThrHisAsnLeuSerSerAsnProSerThrProLeuSer | 880 |
| QY | 4353 | CATCATCCCCCAAGGAATTTGAGTCTTTCGGTCACTCTCTGCTGGTATGTTAATTG | 4412 |
| | | | |
| Db | 881 | HisLeuSerProLysGlnPheGlnSerPheGlyGlnSerSer----- | 894 |
| QY | 4413 | TTTAAATTTTTTAAAGACAATCAATATGAATTCGGTTAATATAGTTATATATTACA | 4472 |

where does it
5482-5601
enc aa 1140-1179?

Db 894 ----- 894
QY 4473 TAACTCGAATAATTGATAGAAAAAATCAGGAATAGAAAAAATAATTATTTCCGACCG 4532
Db 894 ----- 894
QY 4533 CCCATCCATTCTTGAATCCAAATTTGAGAGTATTGTTAGAGATACTACTATTAAA 4592
Db 894 ----- 894
QY 4593 TTTAAACAGCAAAATTCATATCCGTTAATTGAAAAATCACTATTGTTAATAAGAAATTAA 4652
Db 894 ----- 894
QY 4653 AATATGTTTATTATATATTCTTACAGGTGATTAACATGAAAAAGTAGCGCAAGCCAC 4712
Db 895 ----- -Ala--GlyAspAsnMetLysSerArgProSerProG 907
QY 4713 AGGTCACGGGTCAACAGTAATAGTCTAATAGAGGCAAAATTAAGATGTAAGATTGCTG 4772
Db 907 InGlyInArgSerProValAsnSerLeuIleGluAlaAsnLysAspValArgPheAla 927
QY 4773 CATCCAGTCTGTTTAAACCCGACATCCATATGCAAGCAATTCAAATTCAGCATTA 4832
Db 927 IAspSerProGlyPheAsnProHisProHisMetGlnSerAsnSerAlaLeu 947
QY 4833 ACCGCTATAAATGGGCTCTACCAATATACAGATGAGTAATATTAAATATTATTAT 4892
Db 947 snAlaTyrlsMetGlySerThrAsnIleGlnMetGlu--Arg----- 960
QY 4893 TAAAGTTTGTGTTAATTATCTTTTTCAGCGTCAGCATCAGCGCAAGTGATC 4952
Db 961 ----- -GlnAlaSerAlaGlnGlyLys 968
QY 4953 CGTACATTTAGTCGGCGCTCCGATAATATTCGCTAAATCCCAATAGTGCAATCGGCC 5012
Db 968 rValGlnPheSerArgArgSerAspAsnIleProLeuAsnProAsnSerGlyAsnArgPr 988
QY 5013 GCCACCAACAAGATGACCCCAAACTTCGATCCATCTCTTTGGCACAATGTCCCA 5072
Db 988 oProProAsnLysMetThrGlnAsnPheAspProIleSerSerLeuAlaGlnMetSerG 1008
QY 5073 ACAACTAACAAGTTGCGTGTCCAGCATGGGTAGTCCAGCCGAACTGGGTATGACGAT 5132
Db 1008 nGlnLeuThrSerCysValSerSerMetCysSerProAlaGlyThrGlyGlyMetThrMe 1028
QY 5133 GATGGGGGTCGGGACCGTCCGACATCAATATTTGAGCATGGAATATTTCCGGACTAGA 5192
Db 1028 tMetGlyGlyProGlyProSerAspIleAsnIleGlnHisGlyIleIleSerGlyLeuAs 1048
QY 5193 TGGATCAGGAATAGATACCATTAATCAAAATACTGTCAATGATGATGCTAATGAA 5252
Db 1048 pGlySerGlyIleAspThrIleAsnGlnAsnAsnCysHisSerMetAsnValValMetAs 1068
QY 5253 CTCAATGGGTCCCGAATGCTGAATCTTAAATGTGCGTAGCAGGCGTCCAATGAGCC 5312
Db 1068 nSerMetGlyProArgMetLeuAsnProLysMetCysValAlaGlyGlyProAsnGlyPr 1088
QY 5313 GCCTGGCTTTAATCTTAATTTCCCAATGGTGAATTAAGAGAGAATTCATAGGCTG 5372
Db 1088 oProGlyPheAsnProAsnSerProAsnGlyGlyLeuArgGluAsnSerIleGlySerG 1108
QY 5373 CTGTGCTCAGCAAACTCTTCAAACTTTCAAGGGGTGTTCCACCTGGTCCAGATGAT 5432
Db 1108 yCysGlySerAlaAsnSerSerAsnPheGlnGlyValValProProGlyAlaArgMetMe 1128
QY 5433 GGGTCGAATGCCAGTCAATTTTGGTTGCAATTTCAATCCGAATATTCAGGTAAAAGCGAG 5492
Db 1128 tGlyArgMetProValAsnPheGlySerAsnPheAsnProAsnIle----- 1143
QY 5493 TACCCCAACACCATATACATACATGCGAGTAAGGGCACAGAACGCCAACAAATAACAA 5552

Db 1143 ----- 1143
QY 5553 CAATGAGCTAATATATGTGCGAATGCCACCTAGTCTGGAATTTTGCAGAGGTACGCTAA 5612
Db 1144 ----- -GlnArgTyrlaAs 1148
QY 5613 CCTCAATGGGTGCTGTAGGCAATGGGTGCGCAATATGCCACATCAGCCAGCGACGG 5672
Db 1148 nProGlnMetGlyAlaValGlyAsnGlySerProIleCysProProSerAlaSerAspG 1168
QY 5673 TACTCTGGAATGCCAGATTTGATGCGCGGACCAAGAGCGGATGCTAATGAATTC 5732
Db 1168 yThrProGlyMetProGlyLeuMetAlaGlyProGlyAlaGlyGlyMetLeuMetAsnSe 1188
QY 5733 TTCGGAGAGCAACACAGACAAGATCACAAACAATCTGGGGCAAGCATGATTA 5792
Db 1188 rSerGlyGlnGlnHisGlnAsnLysIleThrAsnAsnProGlyAlaSerAsnGlyIleAs 1208
QY 5793 CTTCCTTCAGAAATTCGAATCAATGCTCTATTGTTGACGAAGAGGTGATTACCCGCCA 5852
Db 1208 nPhePheGlnAsnCysAsnGlnMetSerIleValAspGlnGlnGlyLeuProGlyHis 1228
QY 5853 TGACCGATCAATGAATATTGGTCAACCATCTATGATAAGGGGATGCGTCCACATGCCAT 5912
Db 1228 sAspGlySerMetAsnIleGlyGlnProSerMetIleArgGlyMetArgProHisAlaMe 1248
QY 5913 GCGGCCAAATGTAATGGGTGCGCGGATGCCACCCGTTAACAGGCAATTCAGTTGCACA 5972
Db 1248 tArgProAsnValMetGlyAlaArgMetProProValAsnArgGlnIleGlnPheAlaG 1268
QY 5973 GTCATCGGATGTTATGACTGTGTGCGGGAATCCGTCATCATTTTCACTAACGCTTCCTG 6032
Db 1268 nSerSerAspGlyIleAspCysValGlyAspProSerSerPhePheThrAsnAlaSerCy 1288
QY 6033 CAACAGCGCTGACACACACATGTTGGATCAGACACAACAGGCCAATCAGCCTAAGACACA 6092
Db 1288 sAsnSerAlaGlyProHisMetPheGlySerAlaGlnGlnAlaAsnGlnProLysThrG 1308
QY 6093 ACACATTAAGAATACCTAGTGAATGTGTCAAAACCAATCCGGACTTGCAAGTGGCACA 6152
Db 1308 nHisIleLysAsnIleProSerGlyMetCysGlnAsnGlnSerGlyLeuAlaValAlaG 1328
QY 6153 AGGCGAGATCCAATGTCATGGGCAAGACATGCGCAGGGTCACTTTAATTGACCTAC 6212
Db 1328 nGlyGlnIleGlnLeuHisGlyGlnGlnHisAlaGlnGlnGlnSerLeuIleGlyProth 1348
QY 6213 TAAATAAATTTAATGTCAACTGCGGGAAGTGTCACTACTAACGGTCTCTGGCAT 6272
Db 1348 rAsnAsnAsnLeuMetSerThrAlaGlySerValSerAlaThrAsnGlyValSerGlyI 1368
QY 6273 CAATTTCGTAGTCCCTCTTCTACGAGCCTGAAGTATGCCCAATATCATAGTTTCA 6332
Db 1368 eAsnPheValGlyProSerSerThrAspLeuLysTyrlaGlnGlnIntyrHisSerPheG 1388
QY 6333 GCAGAGTTATATGCTACCAACACAGAGAAGTCAACACACAGCATATGACACAGACA 6392
Db 1388 nGlnGlnLeuTyrlaThrAsnThrArgSerGlnGlnGlnHisMetHisGlnGlnHis 1408
QY 6393 CCAGAGCAACATGATAACAATGCCGCCGAATTTATCACCAAAATCCAACTTCTTGTCA 6452
Db 1408 sGlnSerAsnMetIleThrMetProProAsnLeuSerProAsnProThrPhePheValAs 1428
QY 6453 CAAA 6456
Db 1428 nLys 1429

RESULT 2
US-09-915-543-10
; Sequence 10, Application US/09915543
; Publication No. US20020086986A1
; GENERAL INFORMATION:
; APPLICANT: BASLER, Konrad
; APPLICANT: BRUNNER, Erich


```
; APPLICANT: FROESCH, Barbara
; APPLICANT: KRAMPS, Thomas
; APPLICANT: PETER, Oliver
; TITLE OF INVENTION: ESSENTIAL DOWNSTREAM COMPONENT OF THE WINGLESS SIGNALING PATHWAY
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC APPLICATIONS BASED THEREON
; FILE REFERENCE: Q60361
; CURRENT APPLICATION NUMBER: US/09/915,543
; CURRENT FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: 60/221,502
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 112
; TYPE: PRT
; ORGANISM: Drosophila 1gs
US-09-915-543-10

Alignment Scores:
Pred. No.:      2e-38      Length:      112
Score:          570.00     Matches:      112
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match:    4.74%     Indels:      0
DB:             9         Gaps:        0

US-10-664-859-1 (1-6909) x US-09-915-543-10 (1-112)

QY      4029 TCTGCGTCAAGTACCAATAGCCACTCAATCGCCCAATCCCTCGAGTCCAAACAATCTATCT 4088
Db      1 SerAlaSerValProIleAlaThrGlnSerProAsnProSerSerProAsnAsnLeuSer 20

QY      4089 CTCCTCGTACCCGCGGACCAACCGCAGCAGTCAATGGGATTGCCGACCAACTCTCTTAGCATG 4148
Db      21 LeuProSerProArgThrThrAlaAlaValMetGlyLeuProThrAsnSerProSerMet 40

QY      4149 GATGAACAGAGATCATTTATCTGGATCTGTTCCGCAAGCTAATACTTCGACGGTTCAGGCA 4208
Db      41 AspGlyThrGlySerLeuSerGlySerValProGlnAlaAsnThrSerThrValGlnAla 60

QY      4209 GGCACAACAAGTGTCTCTCAGCAACAAGACTGTTTTCAGGACAGACCCCATCGCCG 4268
Db      61 GlyThrThrThrValLeuSerAlaAsnLysAsnCysPheGlnAlaAspThrProSerPro 80

QY      4269 TCAATCAAAATCGTAGTAAGAATACCGGATCGTCAAGCGTTCTTAGCGATACTTAAGC 4328
Db      81 SerAsnGlnAsnArgSerArgAsnThrGlySerSerSerValLeuThrHisAsnLeuSer 100

QY      4329 AGCAACCCCAAGTACCCCTTATCTCATCTATCCCA 4364
Db      101 SerAsnProSerThrProLeuSerHisLeuSerPro 112

RESULT 3
US-10-322-579-10
; Sequence 10, Application US/10322579
; Publication No. US20030114413A1
; GENERAL INFORMATION:
; APPLICANT: BASLER, Konrad
; APPLICANT: BRUNNER, Erich
; APPLICANT: FROESCH, Barbara
; APPLICANT: KRAMPS, Thomas
; APPLICANT: PETER, Oliver
; TITLE OF INVENTION: ESSENTIAL DOWNSTREAM COMPONENT OF THE WINGLESS SIGNALING PATHWAY
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC APPLICATIONS BASED THEREON
; FILE REFERENCE: Q60361
; CURRENT APPLICATION NUMBER: US/10/322,579
; CURRENT FILING DATE: 2002-12-19
; PRIOR APPLICATION NUMBER: US/09/915,543
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: 60/221,502
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.1
```

```
; SEQ ID NO 10
; LENGTH: 112
; TYPE: PRT
; ORGANISM: Drosophila 1gs
US-10-322-579-10

Alignment Scores:
Pred. No.:      2e-38      Length:      112
Score:          570.00     Matches:      112
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match:    4.74%     Indels:      0
DB:             14        Gaps:        0

US-10-664-859-1 (1-6909) x US-10-322-579-10 (1-112)

QY      4029 TCTGCGTCAAGTACCAATAGCCACTCAATCGCCCAATCCCTCGAGTCCAAACAATCTATCT 4088
Db      1 SerAlaSerValProIleAlaThrGlnSerProAsnProSerSerProAsnAsnLeuSer 20

QY      4089 CTCCTCGTACCCGCGGACCAACCGCAGCAGTCAATGGGATTGCCGACCAACTCTCTTAGCATG 4148
Db      21 LeuProSerProArgThrThrAlaAlaValMetGlyLeuProThrAsnSerProSerMet 40

QY      4149 GATGAACAGAGATCATTTATCTGGATCTGTTCCGCAAGCTAATACTTCGACGGTTCAGGCA 4208
Db      41 AspGlyThrGlySerLeuSerGlySerValProGlnAlaAsnThrSerThrValGlnAla 60

QY      4209 GGCACAACAAGTGTCTCTCAGCAACAAGACTGTTTTCAGGACAGACCCCATCGCCG 4268
Db      61 GlyThrThrThrValLeuSerAlaAsnLysAsnCysPheGlnAlaAspThrProSerPro 80

QY      4269 TCAATCAAAATCGTAGTAAGAATACCGGATCGTCAAGCGTTCTTAGCGATACTTAAGC 4328
Db      81 SerAsnGlnAsnArgSerArgAsnThrGlySerSerSerValLeuThrHisAsnLeuSer 100

QY      4329 AGCAACCCCAAGTACCCCTTATCTCATCTATCCCA 4364
Db      101 SerAsnProSerThrProLeuSerHisLeuSerPro 112

RESULT 4
US-10-664-859-10
; Sequence 10, Application US/10664859
; Publication No. US20040038901A1
; GENERAL INFORMATION:
; APPLICANT: BASLER, Konrad
; APPLICANT: BRUNNER, Erich
; APPLICANT: FROESCH, Barbara
; APPLICANT: KRAMPS, Thomas
; APPLICANT: PETER, Oliver
; TITLE OF INVENTION: ESSENTIAL DOWNSTREAM COMPONENT OF THE WINGLESS SIGNALING PATHWAY
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC APPLICATIONS BASED THEREON
; FILE REFERENCE: Q60361
; CURRENT APPLICATION NUMBER: US/10/664,859
; CURRENT FILING DATE: 2003-09-22
; PRIOR APPLICATION NUMBER: US/09/915,543
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: 60/221,502
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 112
; TYPE: PRT
; ORGANISM: Drosophila 1gs
US-10-664-859-10

Alignment Scores:
Pred. No.:      2e-38      Length:      112
Score:          570.00     Matches:      112
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match:    4.74%     Indels:      0
```

| DB: | 15 | Gaps: | 0 |
|-----------------|----------|--|---------|
| US-10-664-859-1 | (1-6909) | x US-10-664-859-10 | (1-112) |
| QY | 4029 | TCTGCGTCAGTACCAATAGCCACTCAATCGCCCAATCCCTCGAGTCCAAACAATTTACT | 4088 |
| Db | 1 | SerAlaSerValProIleAlaThrGlnSerProAsnProSerSerProAsnAsnLeuSer | 20 |
| QY | 4089 | CTCCCGTCACCGCGGACAACCGCAGCAGTCATGGATTGCCAACCACTCTCTTAGCATG | 4148 |
| Db | 21 | LeuProSerProArgThrThrAlaAlaValMetGlyLeuProThrAsnSerProSerMet | 40 |
| QY | 4149 | GATGAACAGGATCATTTACTGATCTGTTCCGCAAGCTAATACTTGACGGTTTCAGCA | 4208 |
| Db | 41 | AspGlyThrGlySerLeuSerGlySerValProGlnAlaAsnThrSerThrValGlnAla | 60 |
| QY | 4209 | GGCACAACAACAGTGTCTTCAGCAACAAGAAGTGTTCAGGCAGACACCCCATCGCG | 4268 |
| Db | 61 | GlyThrThrThrValLeuSerAlaAsnLysAsnCySPhGlnAlaAspThrProSerPro | 80 |
| QY | 4269 | TCAAATCAAATCGTAGTAGAATAACCGGATCGTCAAGCGTTCTTACGCATACTTAAGC | 4328 |
| Db | 81 | SerAsnGlnAsnArgSerArgAsnThrGlySerSerSerValLeuThrHisAsnLeuSer | 100 |
| QY | 4329 | AGCAACCCCAAGTACCCCTTATCTCATCTTATCCCCA | 4364 |
| Db | 101 | SerAsnProSerThrProLeuSerHisLeuSerPro | 112 |

```

RESULT 5
US-10-276-774-2178
; Sequence 2178, Application US/10276774
; Publication No. US20040053245A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; APPLICANT: Tang, Y, Tom et al
; TITLE OF INVENTION: NO. US20040053245A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-030
; CURRENT APPLICATION NUMBER: US/10/276,774
; CURRENT FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 2700
; SOFTWARE: Custom
; SEQ ID NO 2178
; LENGTH: 1435
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-276-774-2178

```

| | |
|------------------------|---------------|
| Alignment Scores: | |
| Pred. No.: | 3.14e-17 |
| Score: | 321.50 |
| Percent Similarity: | 30.10% |
| Best Local Similarity: | 19.11% |
| Query Match: | 2.67% |
| DB: | 15 |
| | Gaps: |
| | 71 |
| | Length: |
| | 1435 |
| | Matches: |
| | 313 |
| | Conservative: |
| | 180 |
| | Mismatches: |
| | 522 |
| | Indels: |
| | 624 |

US-10-664-859-1 (1-6909) X US-10-276-774-2178 (1-1435)

| | | | |
|----|------|--|------|
| QY | 2412 | TCAAACGACAGTACCACGGAAAAAAGGATCGCTCTTGACAATGAATATGACGAAATG | 2471 |
| | ::: | ::: ::: | |
| Db | 55 | AsnThrGlnSerSerProLysSerLysGlnGluValMetValArgProProThrValMet | 74 |
| QY | 2472 | AGCATGGAAGGCTGCAATCAGTTGAATCCGATTTTATCAATGAATCTTTAAATAATCCT | 2531 |
| | | ::: ::: | |
| Db | 75 | SerProSerGlyAsnProGlnLeuAspSerLysPheSerAsnGlnGlyLysGlnGlyGly | 94 |
| QY | 2532 | GCAATTCGACGACATATTAGTAAGC----- | 2555 |
| | ::: | | |
| Db | 95 | SerAlaSerGlnSerGlnProSerProCysAspSerLysSerGlyGlyHisThrProLys | 114 |

| | | | |
|----|------|---|------|
| QY | 2556 | -----GGAGTAGGACCAATAACCCCGAATCGGAATTGGAGCGGGGACGGGAATTTTA | 2606 |
| Db | 115 | AlaLeuProGlyPProGlyGIySerMetGlyLeuLysAsnGlyAlaGlyAsnGly---- | 132 |
| QY | 2607 | TTGACTGCCAACGCCAATGGA-----ATCTCCTCGGGTAGC----- | 2642 |
| Db | 133 | -----AlaLysGlyLySGlyLysArgGluArgSerIleSeraLaAspSerPheAspGln | 150 |
| QY | 2643 | -----AGTAAATTGTGGATTAC | 2660 |
| Db | 151 | ArgAspProGlyThrProAsnAspAspSerAspIleLysGluCyAsnSeraLaAspHis | 170 |
| QY | 2661 | ATGCAACAGCAAATTCAC----- | 2678 |
| Db | 171 | IleLysSerGlnAspSerGlnHisThrProHisSerMetThrProSerAsnAlaThrAla | 190 |
| QY | 2678 | ----- | 2678 |
| Db | 191 | ProArgSerSerThrProProHisGlyGlnThrThrAlaThrGluProThrProAlaGln | 210 |
| QY | 2679 | -----ATATTCGTGTTTCCACTCAGCTGGCCCAACAAGGGCCGAA | 2720 |
| Db | 211 | LysThrProAlaLysValValTyrrValPheSerThrGluMetAlaAsnLysAlaAlaGlu | 230 |
| QY | 2721 | TCAGTTTTTAAGCGGTCAATTTCAAACTATTATTCGTTACTGTCACTGCAGCTGTCTACA | 2780 |
| Db | 231 | AlaValLeuLysGlyGlnValGluThrIleValSerPheHis----- | 244 |
| QY | 2781 | AAAAGCTTCCTGGAAGACTTTTATGAAAAACCCTTTAAGATTAAAGATTACAGCGG | 2840 |
| Db | 245 | -----IleGlnAsn--IleSerAsnAsnLysThrGluArg | 255 |
| QY | 2841 | CACAATTCGTCGGTGATGCCATGATAGGCATGGGGCAGTTGGACTAACTCTCCTTAAT | 2900 |
| Db | 256 | SerThrAla-----ProLeuAsnThrGlnIleSeraLaLeuArgAsnAspProLys | 272 |
| QY | 2901 | CCTGTAGCCNAATAACACACAGCAGCCCATACAAAGACCCTATTTGAAAACC | 2960 |
| Db | 273 | ProLeu-----ProGlnGlnProProAlaProAla----- | 282 |
| QY | 2961 | CAATTCAATCAACATGAAAAACAGCAACCGTAGTACTGTAAGCGCGCTAGCAACTCTTTT | 3020 |
| Db | 283 | -----AsnGlnAspGlnAsnSerSerGlnAsnThrArgLeuGlnProThrProProIle | 300 |
| QY | 3021 | -----GTGCACCAGTCTGATCTATG | 3041 |
| Db | 301 | ProAlaProAlaProLysProAlaAlaProProAlaArgProLeuAspArgLysSerPro--- | 319 |
| QY | 3042 | GGCAACGAACTGAATTGATGTGCTGGGAAGCGCGATCTCTCAAAACACCACT----- | 3092 |
| Db | 320 | GlyValGlnAsnLysLeuIleProSerValGlySerProAlaSerSerThrProLeuPro | 339 |
| QY | 3093 | -----AGGTCGTGGACAAACTCACGAATCATGTAGACAGTATCATGTACATCCACGGAG | 3146 |
| Db | 340 | ProAspGlyThrGlyProAsnSerThrProAsnAsnArgAlaValThrProValSerGln | 359 |
| QY | 3147 | TCACAGGCAATAAGATACTGGAAGCAGCTGGCGTTGATTGGGACAGGTTCACAAAAAGGA | 3206 |
| Db | 360 | GlySerAsnSerSerSeraLaAspProLysAlaProProProProValSerSergly | 379 |
| QY | 3207 | AGGATCCTGGCCTGACAACTGAAAAACAATGTATCACTGCAAGGATTAAAGTTCCA | 3266 |
| Db | 380 | GluProProThrLeu--GlyGluAsn-----Pro | 388 |
| QY | 3267 | GACGAAAAACCTTACACACACAGCGGCAACATCGGGAAGAACAGTTGGCAAAAATTAATAA | 3326 |
| Db | 389 | Asp---GlyLeuSerGlnGluGlnLeuGlnHisArgGluArgSerLeuGlnThrLeuArg | 407 |
| QY | 3327 | AAATGAATCAATTTCTTTTCTCGAAATGAGAATTCAGTAGAGCTAATGTAAGCTCA | 3386 |
| Db | 408 | AspIleGlnArgMetLeuPheProAspGluLysGluPheThrGlyAlaGln----- | 424 |
| QY | 3387 | CAGATAACAAAAATTTCCAGGAGATTTAATGATGGGGATGTCCGGTGGCGGACCGGATCT | 3446 |

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: November 2, 2005, 21:47:47 ; Search time 164 Seconds
(without alignments)
6289.639 Million cell updates/sec

Title: US-10-664-859-1
Perfect score: 12037
Sequence: 1 acgagtgctctctcttattat.....aaaaaaaaaaaaaaaaaaaaa 6909

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 1027090

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/p/US10664859/runat_02112005_171804_3553/app_query.fasta_1.7047
-DB=Issued_Patents_AA -QFMT=faastan -SUFFIX=ra1 -MINMATCH=0.1 -LOOPEXT=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=biosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10664859@cgn_1.181@runat_02112005_171804_3553 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/5A.COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/5B.COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PTUS.COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|---------------------------------------|
| 1 | 2429.5 | 20.2 | 506 | 4 | US-09-270-767-44797 Sequence 44797, A |
| 2 | 1573 | 13.1 | 317 | 4 | US-09-270-767-60260 Sequence 60260, A |
| 3 | 1145 | 9.5 | 259 | 4 | US-09-270-767-46113 Sequence 46113, A |
| 4 | 935 | 7.8 | 175 | 4 | US-09-270-767-36663 Sequence 36663, A |
| 5 | 935 | 7.8 | 175 | 4 | US-09-270-767-51880 Sequence 51880, A |
| 6 | 599.5 | 5.0 | 230 | 4 | US-09-270-767-44999 Sequence 44999, A |
| 7 | 594 | 4.9 | 140 | 4 | US-09-270-767-36764 Sequence 36764, A |
| 8 | 594 | 4.9 | 140 | 4 | US-09-270-767-51981 Sequence 51981, A |
| 9 | 312 | 2.6 | 65 | 4 | US-09-270-767-43686 Sequence 43686, A |
| 10 | 238.5 | 2.0 | 1179 | 4 | US-09-949-016-7088 Sequence 7088, Ap |
| 11 | 236 | 2.0 | 43 | 4 | US-09-270-767-61668 Sequence 61668, A |
| 12 | 227 | 1.9 | 1076 | 4 | US-09-949-016-7421 Sequence 7421, Ap |

| | | | | | | |
|----|-------|-----|------|---|----------------------|--------------------|
| 13 | 222 | 1.8 | 1420 | 4 | US-09-125-635-4 | Sequence 4, Appli |
| 14 | 219 | 1.8 | 2451 | 4 | US-09-949-016-9675 | Sequence 9675, Ap |
| 15 | 216.5 | 1.8 | 1061 | 3 | US-08-701-154A-5 | Sequence 5, Appli |
| 16 | 204 | 1.7 | 1584 | 3 | US-09-457-040B-27 | Sequence 27, Appli |
| 17 | 203 | 1.7 | 2314 | 3 | US-09-268-347-49 | Sequence 49, Appli |
| 18 | 202 | 1.7 | 1600 | 2 | US-08-617-697-10 | Sequence 10, Appli |
| 19 | 201 | 1.7 | 1036 | 3 | US-08-891-640-3 | Sequence 3, Appli |
| 20 | 201 | 1.7 | 1302 | 4 | US-09-949-016-10852 | Sequence 10852, A |
| 21 | 198.5 | 1.6 | 1258 | 4 | US-09-538-092-886 | Sequence 886, App |
| 22 | 198.5 | 1.6 | 1531 | 4 | US-09-949-016-6740 | Sequence 6740, App |
| 23 | 198 | 1.6 | 2703 | 1 | US-08-185-432-19 | Sequence 19, Appli |
| 24 | 198 | 1.6 | 2703 | 4 | US-08-899-232-4 | Sequence 4, Appli |
| 25 | 198 | 1.6 | 2703 | 4 | US-09-121-457-4 | Sequence 5821, Ap |
| 26 | 196.5 | 1.6 | 2504 | 4 | US-09-328-352-5821 | Sequence 5821, Ap |
| 27 | 194.5 | 1.6 | 1529 | 2 | US-08-728-470-10 | Sequence 10, Appli |
| 28 | 194.5 | 1.6 | 1529 | 3 | US-08-719-641-10 | Sequence 10, Appli |
| 29 | 193 | 1.6 | 908 | 4 | US-09-949-016-8542 | Sequence 8542, Ap |
| 30 | 193 | 1.6 | 2169 | 4 | US-09-252-991A-30227 | Sequence 6930, Ap |
| 31 | 192 | 1.6 | 2736 | 4 | US-09-949-016-6930 | Sequence 30227, A |
| 32 | 191 | 1.6 | 3969 | 3 | US-08-061-376-5 | Sequence 5, Appli |
| 33 | 191 | 1.6 | 3969 | 4 | US-09-538-092-1262 | Sequence 1262, Ap |
| 34 | 189 | 1.6 | 1268 | 4 | US-09-270-767-45320 | Sequence 45320, A |
| 35 | 186.5 | 1.5 | 1776 | 4 | US-09-556-877-179 | Sequence 179, App |
| 36 | 186.5 | 1.5 | 1776 | 4 | US-09-620-412C-179 | Sequence 179, App |
| 37 | 186.5 | 1.5 | 1776 | 4 | US-09-598-419-179 | Sequence 179, App |
| 38 | 183.5 | 1.5 | 1187 | 4 | US-09-598-419-179 | Sequence 6513, Ap |
| 39 | 183.5 | 1.5 | 1402 | 4 | US-09-445-353E-2 | Sequence 2, Appli |
| 40 | 181.5 | 1.5 | 1477 | 4 | US-09-206-942-71 | Sequence 71, Appli |
| 41 | 180.5 | 1.5 | 1992 | 4 | US-08-621-944A-3 | Sequence 3, Appli |
| 42 | 180.5 | 1.5 | 1992 | 4 | US-08-945-567D-3 | Sequence 3, Appli |
| 43 | 179.5 | 1.5 | 1833 | 4 | US-08-621-944A-4 | Sequence 4, Appli |
| 44 | 179.5 | 1.5 | 1833 | 4 | US-08-945-567D-4 | Sequence 4, Appli |
| 45 | 176.5 | 1.5 | 1323 | 4 | US-09-248-796A-19543 | Sequence 19543, A |

ALIGNMENTS

RESULT 1
US-09-270-767-44797
Sequence 44797, Application US/09270767
Patent No. 6703491
GENERAL INFORMATION:
APPLICANT: Homburger et al.
TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
FILE REFERENCE: File Reference: 7326-094
CURRENT APPLICATION NUMBER: US/09/270,767
CURRENT FILING DATE: 1999-03-17
NUMBER OF SEQ ID NOS: 62517
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 44797
LENGTH: 506
TYPE: PRT
ORGANISM: Drosophila melanogaster
FEATURE:
OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-44797

Alignment Scores:
Pred. No.: 6.03e-220
Score: 2429.50
Percent Similarity: 80.03%
Best Local Similarity: 79.71%
Query Match: 20.18%
DB: 4
Gaps: 2

US-10-664-859-1 (1-6909) x US-09-270-767-44797 (1-506)
QY 3271 AAACCTTACACACAGCGGACATCGGAGAGACAGTGGCAAAATATAAAAA 3330
Db 1 LysThrLeuHisHisAsnSerGlyAsnIleGlyLysAsnSerTrpGlnLys**LysLys 20
QY 3331 TGAATCAATTTCTTTCTTCTGAAATGAGATTAGTAGAGCTAATGTAGCTACAGA 3390
|||||

Ag data. uspt. gov
Ag data. gov
Pat ID 6703491 B1

Db 21 ***IleAsnPhePheLeuLysMetArgIleGln**GluLeuMet**AlaHisArg 40
Qy 3391 TAACAAAAATTCAGAGATTATGATGGGATGTCGGGTGGC-EGAGCGGATCTATT 3449
Db 41 ***GlnPheGlnGlnIle****TrrGlyCysGlnValAla^{91b}GlyGlySerIle 60
Qy 3450 ATAATCCGACGATGCGACAACTGCATATGCGAGTAAAGCCAAATCGAGCTTTATCG 3509
Db 61 IleAsnProThrmMetArgGlnLeuHisMetProGlyAsnAlaLysSerGluLeuLeuSer 80
Qy 3510 GCGACAAGTTCAGGACTTTCGAGAGATGTAATGCATCCAGGGAGTGTATATCAGATATG 3569
Db 81 AlaThrSerSerGlyLeuSerGluAspValMetHisProGlyAspValIleSerAspMet 100
Qy 3570 GGTGCGGTATAGATGTAATATAATCAAAAAACAGTGTGCAATGTGATCTGAGTA 3629
Db 101 GlyAlaValIleGlyCysAsnAsnGlnLysThrSerValGlnCysGlySerGlyVal 120
Qy 3630 GGTGTTGCTCACTGGAACAACACTGCAGCTGGAGTAAATGTCATATGATGCTCAAGCTCC 3689
Db 121 GlyValValThrGlyThrThrAlaIleGlyValAsnValAsnMetHisCysSerSerSer 140
Qy 3690 GGGCGCCCGAATGGCAATATGATGGAGCTCTACGATATGCTAGCCTCGTTGGCAAC 3749
Db 141 GlyAlaProAsnGlyAsnMetMetGlySerSerThrAspMetLeuAlaSerPheGlyAsn 160
Qy 3750 ACAAGCTGCAACGTTCATCGAAGCGGCCAGATATGTTCTAAGGAAGTTTAAATCAAGAT 3809
Db 161 ThrSerCysAsnValIleGlyThrAlaProAspMetSerLysGluValLeuAsnGlnAsp 180
Qy 3810 AGCCGAACCCATTCATCAATGAGGGAGTTCATGAGTTCGAGTTCGAGTTCGAGTTC 3869
Db 181 SerArgThrHisSerHisGlnGlyGlyValAlaGlnMetGluTrpSerLysIleGlnHis 200
Qy 3870 CAATTTTTCGAGAAGCGCTTCAAGGGGGGAGCCAGACAAGTCACTGGAATCTAGTA 3929
Db 201 GlnPhePheGlnGluArgLeuLysGlyLysProArgGlnValThrGlyThrValVal 220
Qy 3930 CCACACACAGCAAAACCCCTTCTGATCTGTGGAACCTGTTAAACAACAGGTGCGACC 3989
Db 221 ProGlnGlnGlnThrProSerGlySerGlyGlyAsnSerLeuAsnAsnGlnValArgPro 240
Qy 3990 CTGCAAGGTCCACCTCTCTCTTACCACTCCATCCAGAGATCTGCGTCACTACCAATAGCC 4049
Db 241 LeuGlyGlyProProProProTyHisSerIleGlnArgSerAlaSerValProIleAla 260
Qy 4050 ACTCAATCGCCCAATCCCTCGAGTCCAAACAATCTATCTCTCCCGTCAACCGGACAACC 4109
Db 261 ThrGlnSerProAsnProSerSerProAsnAsnLeuSerLeuProSerProArgThrThr 280
Qy 4110 GCAGCAGTCATGGATGTCGCGACCAACTCTCTAGCATGATGGAACAGATCATTTATCT 4169
Db 281 AlaAlaValMetGlyLeuProThrAsnSerProSerMetAspGlyThrGlySerLeuSer 300
Qy 4170 GGATCTGTTCCGACGCTAATACTTGAAGGTTCAAGGACAGCAACAACAGTGTCTCA 4229
Db 301 GlySerValProGlnAlaAsnThrSerThrValGlnAlaGlyThrThrValLeuSer 320
Qy 4230 GCAACAACAAGACTGTTTTCAGGACAGACACCCCACTGCCGTCAAATCAAAATCGTAGTA 4289
Db 321 AlaAsnLysAsnCybPheGlnAlaAspThrProSerProSerAsnGlnAsnArgSerArg 340
Qy 4290 AATACCGGATCGTCAAGCGTCTTAGCGATTAATTAAGCAGCAACCCAAAGTACCCCTTA 4349
Db 341 AsnThrGlySerSerSerValLeuThrHisAsnLeuSerSerAsnProSerThrProLeu 360
Qy 4350 TCTCATCTATCCCAAGGAATTGAGTCTTTCGTCAGTCTCTGCTGCTATGTTATAT 4409
Db 361 SerHisLeuSerProLysGlnPheGlnSerPheGlyGlnSerSer 375
Qy 4410 TTGTTTAAATTTTAAAGACAATAATCAAAATATGAATTGCGTTAATAATAGTTATATAT 4469
Db 375 ----- 375

Qy 4470 ACATAACTCGAAATTTGATAGAAAAAATCAGGAATAGAAAAAATAATTATTTCCGGA 4529
Db 375 ----- 375
Qy 4530 CCGCCCATCCATTTCTGAATCCAATTTCTGAGTGAATTGTTAGAGATAATCTACTATTA 4589
Db 375 ----- 375
Qy 4590 AAATTAACACGAAATTCATATCCGTTAATTGAAAAATCACTATTGTTAATAAGAAATT 4649
Db 375 ----- 375
Qy 4650 AAAAAATATGTTTATTAATATTTCTACAGGTGATTAATGAAAAAGTAGCGCAAGCC 4709
Db 376 -----Ala-GlyAspAsnMetLysSerArgArgProSerP 387
Qy 4710 CACAGGTCAGCGGTACCAAGTAAATAGTCTTAATAGAGCAAAATAAGATGTACGATTTG 4769
Db 387 roGlnGlyGlnArgSerProValAsnSerLeuIleGluAlaAsnLysAspValArgPhea 407
Qy 4770 CTGCATCCAGTCTGTTTAAACCCGCATCCACATATGCAAGCAATTCAATTGACAT 4829
Db 407 IaAlaSerSerProGlyPheAsnProHisProHisMetGlnSerAsnSerAsnSerAlaL 427
Qy 4830 TAAACGCTTATAAATGGGCTCTACCAATATATACAGATGAGAGTAAATATTAATTTT 4889
Db 427 euaenAlaTyrLysMetGlySerThrAsnIleGlnMetGlu--Arg- 441
Qy 4890 ATTAAAGTTTGTGTATTTATTTCTTTTTCAGGCTCAAGCATCAGCGCAAGGTGG 4949
Db 442 -----GlnAlaSerAlaGlnGlyGly 448
Qy 4950 ATCCGTAACAATTTAGTCGGCGCTCCGATATATTTCCGCTAAATCCCAATAGTGCAATCG 5009
Db 448 ySerValGlnPheSerArgArgSerAspAsnIleProLeuAsnProAsnSerGlyAsnAr 468
Qy 5010 GCCGCCACCAACAAGATGACCCAAAACCTTGATCCAACTCTTCTTTGGACAATGTC 5069
Db 468 gProProProAsnLysMetThrGlnAsnPheAspProIleSerSerLeuAlaGlnMetSe 488
Qy 5070 CCAACAACCTAACAAAGTTGCGTTCACAGATGGTAGTCCAGCCGGAATGTGT 5124
Db 488 rGlnGlnLeuThrSerCysValSerSerMetGlySerProAlaGlyThrGlyGly 506

RESULT 2
US-09-270-767-60260 1020
; Sequence 60260, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 60260
; LENGTH: 317
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: xaa means any amino acid
US-09-270-767-60260

Alignment Scores:
Pred. No.: 2.8e-139 Length: 317
Score: 1573.00 Matches: 308
Percent Similarity: 97.48% Conservative: 1
Best Local Similarity: 97.16% Mismatches: 8
Query Match: 13.07% Indels: 1
DB: 4 Gaps: 0

US-10-664-859-1 (1-6909) X US-09-270-767-60260 (1-317)

| | | | |
|----|------|--|------|
| QY | 3271 | AAACCTTACACCAACAGCGGCAACATCGGGAGAGAACAGTTGGCAAAAATAAAAAA | 3330 |
| Db | 1 | LysThrLeuHisHisSerGlyAsnIleGlyLysAsnSerTrpGlnLys**LysLys | 20 |
| QY | 3331 | TGAATCAATTTCTTTTCTGAAATGAGAATTCAAGAGACTAATGTAAAGCTCACAGA | 3390 |
| Db | 21 | **IleAsnPhePheLeuLysMetArgIleGln**GluLeuMet**AlaHisArg | 40 |
| QY | 3391 | TAACAAAAATTCAGAGAGATTAAATGATGGGAGTGCGGTGGC -GAGAGCGATCTAAT | 3449 |
| Db | 41 | **GlnLysPheGlnGluIle*****TrpGlyCysGlnValAlaGlyGlyGlySerIle | 60 |
| QY | 3450 | ATAAATCCGACGATGGCGAACCTGCATATGCCAGGTAAAGCCAAATCGGAGCTTTATCG | 3509 |
| Db | 61 | IleAsnProThrMetArgGlnLeuHisMetProGlyAsnAlaLysSerGluLeuLeuSer | 80 |
| QY | 3510 | GCGACAAGTTCAGAGACTTTCGGAAGATGTATATGCATCCAGGGGATGTTATATCAGATATG | 3569 |
| Db | 81 | AlaThrSerSerGlyLeuSerGlyAspValMetHisProGlyAspValIleSerAspMet | 100 |
| QY | 3570 | GGTGCCGTAATAGGATGTAAATTAATCAAAAAACCAAGTGTCAATGTGATCTGAGATG | 3629 |
| Db | 101 | GlyAlaValIleGlyCysAsnAsnAsnGlnLysThrSerValGlnCysGlySerGlyVal | 120 |
| QY | 3630 | GGTGTGTCTACTGGAACACTGCAGCTGGAGTAAATGTCAATATGCATTTGCTCAAGCTCC | 3689 |
| Db | 121 | GlyValValThrGlyThrThrAlaIaGlyValAsnValAsnMetHisCysSerSerSer | 140 |
| QY | 3690 | GGCGCCCCGAATGGCAATATGATGGGAAGCTCTACGGATATGCTAGCCCTGTTGGCAAC | 3749 |
| Db | 141 | GlyAlaProAsnGlyAsnMetMetGlySerSerThrAspMetLeuAlaSerPheGlyAsn | 160 |
| QY | 3750 | ACAAGCTGCAACGTCAATCGGAACGGCCCCAGATATGTCTAAGGAAGTTTAAATCAAGAT | 3809 |
| Db | 161 | ThrSerCysAsnValIleGlyThrAlaProAspMetSerLysGluValLeuAsnGlnAsp | 180 |
| QY | 3810 | AGCCGAACCCATTCACATCAAGGGGGAGTTGCTCAATGAGAGTGTGGAAGATTCAACAT | 3869 |
| Db | 181 | SerArgThrHisSerHisGlnGlyGlyValAlaGlnMetGluTrpSerLysIleGlnHis | 200 |
| QY | 3870 | CAATTTTTCGAAGAACGCTCTAAGGGGGGCAAGCCAGACAAGTCACTGGAAGTGTAGTA | 3929 |
| Db | 201 | GlnPhePheGlnGluArgLeuLysGlyGlyLysProArgGlnValThrGlyThrValVal | 220 |
| QY | 3930 | CCACAACAGCAAAACCCCTTCTGGATCTGTGTGAAACTGTAAACAACCAAGTGCAGACC | 3989 |
| Db | 221 | ProGlnGlnGlnThrProSerGlySerGlyLysAsnSerLeuAsnAsnGlnValArgPro | 240 |
| QY | 3990 | CTGCAAGGTCCACCTCTCTTACCACTCCATCCAGAGATCTGCGTCAGTACCAATAGCC | 4049 |
| Db | 241 | LeuGlnGlyProProProProTyrHisSerIleGlnArgSerAlaSerValProIleAla | 260 |
| QY | 4050 | ACTCAATCGCCCAATCCCTCGAGTCCAAACAATCTATCTCCCGTCACCGCGACAACC | 4109 |
| Db | 261 | ThrGlnSerProAsnProSerSerProAsnAsnLeuSerLeuProSerProArgThrThr | 280 |
| QY | 4110 | GCAGCAGTCATGGGATTGCCGACCAACTCTCCTAGCATGATGGAACAGATCATATCT | 4169 |
| Db | 281 | AlaAlaValMetGlyLeuProThrAsnSerProSerMetAspGlyThrGlySerLeuSer | 300 |
| QY | 4170 | GGATCTGTTCGCAAGCTAATACTTCGACGGTTCAAGCAGGCAACAACAACA 4220 | |
| Db | 301 | GlySerValProGlnAlaAsnThrSerThrValGlnAlaGlyThrThrThr 317 | |

RESULT 3

US-09-270-767-46113

; Sequence 46113, Application US/09270767

Patent No. 6703491

GENERAL INFORMATION:

APPLICANT: Homburger et al.

1. TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*;

```

; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 46113
; LENGTH: 259
; TYPE: PR1
; ORGANISM: Drosophila melanogaster
US-09-270-767-46113

```

| | | |
|------------------------|---------------|-----|
| Alignment Scores: | | |
| Pred. No.: | 5.83e-99 | 259 |
| Score: | 1145.00 | 250 |
| Percent Similarity: | 50.60% | |
| Best Local Similarity: | 50.40% | |
| Query Match: | 9.51% | 244 |
| DB: | 4 | 1 |
| | | |
| | Length: | |
| | Matches: | |
| | Conservative: | |
| | Mismatches: | |
| | Indels: | |
| | Gaps: | |

US-10-664-859-1 (1-6909) x US-09-270-767-46113 (1-259)

| | | | |
|----|------|--|------|
| QY | 1453 | TCAGAAATAGTTCTAATTAAGCAAAAGGTCAGACAGCTGTGGCGGCTGCCAAACAGST | 1512 |
| Db | 8 | AlaGluIleSerSerAsnLysAlaLysGlyGlnAlaAlaGlyGlyCysGluThrGly | 27 |
| QY | 1513 | TCTACATCCAGTTTGACTGTCAAGGAAGAACCCAGATGCTTAGGCAGTTAGTAAAT | 1572 |
| Db | 28 | SerThrSerSerLeuThrValLysGluGluProThrAspValLeuGlySerLeuValAsn | 47 |
| QY | 1573 | ATGAAAAAAGAGAAGAAAGAAATCATTCGCCAAGATGTCCCTGTGGTTTGGTTCA | 1632 |
| Db | 48 | MetLysLysGluGluArgGluAsnHisSerProThrMetSerProValGlyPheGlySer | 67 |
| QY | 1633 | ATTGTAATGCA CAGGACAACTCCGCTACACCGGGTAAGTTTAAAGAGATCCATATAAG | 1692 |
| Db | 68 | IleGlyAsnAlaGlnAspAsnSerAlaThr----- | 77 |
| QY | 1693 | CAAAATACAGAAATTAATGTCACTTACCATTTTATTGTAGTCAAGAAGACTACTATAG | 1752 |
| Db | 77 | ----- | 77 |
| QY | 1753 | CGATATCTCCTGCCTTTTAATTTTATTAAATTAGGAATACGAATATTTCTAATTGTA | 1812 |
| Db | 77 | ----- | 77 |
| QY | 1813 | AAATAAAATTGATTAATTAAGTAGAATTTAAAAACCTTTGAATTAGACATACCCTTCC | 1872 |
| Db | 77 | ----- | 77 |
| QY | 1873 | AAAAATCAGTAATCATTTGGGAACGAGAGTGTGTCGCCGAAGAGACTACTATAAACCTT | 1932 |
| Db | 77 | ----- | 77 |
| QY | 1933 | TTGAGCTATCTGATACTGCACGCTACTAAAAATGATTAGTTAGAAAAATGGGTGAATT | 1992 |
| Db | 77 | ----- | 77 |
| QY | 1993 | TTGTAGAAGTTTTCATTTTAGAAGAAATGTGATTATTTATTAACCCCTTCAAGCGA | 2052 |
| Db | 77 | ----- | 77 |
| QY | 2053 | ACTACATTTGTTCTACGATATTTTGAAAAACAATGGTTAAGTTGGAAGTGCTATAA | 2112 |
| Db | 77 | ----- | 77 |
| QY | 2113 | AACAGAATTCACGGTTTCAATACTAACACAGGTTTTTGATTATTTGATTAATGAG | 2172 |
| Db | 77 | ----- | 77 |
| QY | 2173 | AAATTATCAGACTTCAGTTAAATGTTTAATTCGATTAAGGTCGACAATCACAGAGAT | 2232 |
| Db | 77 | ----- | 77 |

QY 2233 TTCATTTTGGGTATATATAGAGTCGCTTCACACTCTTGGCGCGCTTCACCAC 2292
Db 77 ----- 77
QY 2293 TACGTGAGTCCGCCCGCAGTGATTATATAGATGATTACGAGTTATTATTTTTA 2352
Db 77 ----- 77
QY 2353 TGGTGATTTTAATAATATCTTATTATTCACTTTACATAGTTAAATTGAAGAATTT 2412
Db 78 -----Pro-ValylsilegIuArgIles 85
QY 2413 CAAACGACAGTACACCGAAAAAAGATCGTCTTGACAATGAATAATGACGAATGA 2472
Db 85 erAsnAspSerThrThrGluylsGlySerSerleuthrMetAsnAsnAspGluMetS 105
QY 2473 GCATGAAGGCTGCAATCAGTTGAATCCCGATTTTATCAATGAATCTTTAAATAATCCTG 2532
Db 105 erMetGluGlyCysAsnGlnLeuAsnProAspPheIleAsnGlnSerleuAsnAsnProA 125
QY 2533 CAATTTCGACATATTAGTAGCGAGTAGACCAATACCCGGAATCGGAGTTGGAGCGG 2592
Db 125 laIleSerSerIleleuValSerGlyValGlyProIleProGlyIleGlyValGlyAlaG 145
QY 2593 GGACGGGAATTTATTGACTGCCAACGCCAATGGAATCTCTCGGGTAGCAGTAATTGTT 2652
Db 145 lyThrGlyAsnLeuLeuThrAlaAsnAlaAsnGlyIleSerSerGlySerSerAsnCysL 165
QY 2653 TGGATTACATGCAACGCAAAATCACATATTGTTTCAACTCAGCTGGCCAAACAAG 2712
Db 165 euAspIlyrMetGlnGlnGlnAsnHisIlePheValPheSerThrGlnLeuAlaAsnLysG 185
QY 2713 GGGCCGAATCAGTTTAAAGCGGTCAATTTTCAACTATTATTGCGTATGCACTGCACTCAGC 2772
Db 185 lyAlaGlnSerValleuSerGlyGlnPheGlnThrIleIleAlaIyHisCysThrGlnP 205
QY 2773 CTGCTACAAAAAGCTTCTCGTAGAGACTTTTATGAAAAACCTTTAAAGATTAAACAAGT 2832
Db 205 roAlaThrIlySerPheleuGluAspPheMetLysAsnProleuLysIleAsnLysL 225
QY 2833 TACAGCGGCACAATCCGTCGGTATGCCATGATAGGCATGGGCGAGGTTGGACTAATC 2892
Db 225 euGlnArghIleAsnSerValGlyMetProIlyIleGlyMetGlyGlnValGlyLeuThrP 245
QY 2893 CTCCTAATCCTGTAGCCAAATAATACACACAGACCCACATACA 2936
Db 245 roProAsnProValAlaIalysIleThrGlnGlnGlnProHisThr 259
RESULT 4
US-09-270-767-36663
; Sequence 36663, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 36663
; LENGTH: 175
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-36663

Alignment Scores: 2.97e-79 Length: 175
Pred. No.: 935.00 Matches: 175
Score: 935.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 7.77%
DB: 4 Gaps: 0

US-10-664-859-1 (1-6909) x US-09-270-767-36663 (1-175)
QY 5725 ATGAATCTTCCGAGAGACAACCAAGATCACAAACAATCCTGGGCAAGCAAT 5784
Db 1 MetAsnSerSerGlyGlnHisGlnAsnLysIleThrAsnAsnProGlyAlaSerAsn 20
QY 5785 GGTATTAACTTCTTTCAGAAATGCAATCAATGTCTATTGTTGACGAAGAGGTGATTA 5844
Db 21 GlyIleAsnPhePheGlnAsnGlnMetSerIleValAspGluGluGlyLeu 40
QY 5845 CCCGCCATGACGGATCAATGAATATTGGTCAACCATCTATGATAAGGGCATGCGTCCA 5904
Db 41 ProGlyHisAspGlySerMetAsnIleGlyGlnProSerMetIleArgGlyMetArgPro 60
QY 5905 CATGCCATGCGGCCAAATGTATGCGTCCGGGATGCCACCCGTTAACAGGCAATTCAG 5964
Db 61 HisAlaMetArgProAsnValMetGlyAlaArgMetProProValAsnArgGlnIleGln 80
QY 5965 TTTGACAGTCATCGGATGGTATTGACTGTGTGGGGATCCGTCATCTTTTCACTTAAC 6024
Db 81 PheAlaGlnSerSerAspGlyIleAspCysValGlyAspProSerSerPhePheThrAsn 100
QY 6025 GCTTCTGCAACAGCGCTGACCAACATGTTTGATCAGACAAACAGGCCAATCAGCCT 6084
Db 101 AlaSerCysAsnSerAlaGlyProHisMetPheGlySerAlaGlnGlnAlaAsnGlnPro 120
QY 6085 AAGACACAACACATAAAGAACAATACCTAGTGAATGTGTCAAAACCAATCGGACTTGCA 6144
Db 121 LysThrGlnHisIleLysAsnIleProSerGlyMetCysGlnAsnGlnSerGlyLeuAla 140
QY 6145 GTGGCAAGAGGCGAGATCCCACTGCATGGGCAAGACATGCCGAGGTCAGTCTTAAT 6204
Db 141 ValAlaGlnGlyGlnIleGlnLeuHisGlyGlnGlyHisAlaGlnGlyGlnSerLeuIle 160
QY 6205 GGACCTAATAATTAATTTATGTCACACTGCCGGAAGTGTCAGT 6249
Db 161 GlyProThrAsnAsnAsnLeuMetSerThrAlaGlySerValSer 175
RESULT 5
US-09-270-767-51880
; Sequence 51880, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 51880
; LENGTH: 175
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-51880

Alignment Scores: 2.97e-79 Length: 175
Pred. No.: 935.00 Matches: 175
Score: 935.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 7.77%
DB: 4 Gaps: 0

US-10-664-859-1 (1-6909) x US-09-270-767-51880 (1-175)
QY 5725 ATGAATCTTCCGAGAGACAACCAAGATCACAAACAATCCTGGGCAAGCAAT 5784
Db 1 MetAsnSerSerGlyGlnHisGlnAsnLysIleThrAsnAsnProGlyAlaSerAsn 20
QY 5785 GGTATTAACTTCTTTCAGAAATGCAATCAATGTCTATTGTTGACGAAGAGGTGATTA 5844
Db 5785 GGTATTAACTTCTTTCAGAAATGCAATCAATGTCTATTGTTGACGAAGAGGTGATTA 5844

Db 21 GlyIleasnphneheglnasnCysasnGlnMetSerIleValaspGluGluGlyLeu 40
QY 5845 CCCGCCATGACGATCAATGATATTGGTCAACCATCTATGATAAGGGCATGCGTCCA 5904
Db 41 ProGlyHisaspGlySerMetAsnIleGlyGlnProSerMetIleargGlyMetArgPro 60
QY 5905 CATGCCATGCGGCCAAATGTAATGGTGGCGGATGCCACCCGTTAAACGCAAAATTCCAG 5964
Db 61 HisAlaMetargProAsnValMetGlyAlaargMetProProValAsnargGlnIleGln 80
QY 5965 TTTCACAGTCATCCGATGGTATGTGACTGTGTGCGGGATCCGTCATCATTTTTCACATAAC 6024
Db 81 PheAlaGlnSerSerAspGlyIleaspCysValGlyAspProSerSerPhePheThrAsn 100
QY 6025 GCTTCCTGCACACAGCGCTGGACACACATGTTGGATCAGCACACACAGGCCAATCAGCCT 6084
Db 101 AlaSerCysAsnSerAlaGlyProHisMetPheGlySerAlaGlnGlnAlaAsnGlnPro 120
QY 6085 AAGACACAACACATAAAGACATACCTAGTGAATGTGTCAAAACCAATCCGGACTTGCA 6144
Db 121 LysThrGlnHisIleIleLysAsnIleProSerGlyMetCysGlnAsnGlnSerGlyLeuAla 140
QY 6145 GTGGCACAAGGCGAGATCCAACTGCATGGCGCAAGACATGCGCAGGTCAGTCTTTAAT 6204
Db 141 ValAlaGlnGlyGlnIleGlnLeuHisGlyGlnGlyHisAlaGlnGlyGlnSerLeuIle 160
QY 6205 GGACCTACTAATAATAATTTAATGTCACTGCCGGAAGTGTCACT 6249
Db 161 GlyProThrAsnAsnAsnLeuMetSerThrAlaGlySerValSer 175

RESULT 6

US-09-270-767-44999
; Sequence 44999, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 44999
; LENGTH: 230
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-44999

Alignment Scores:
Pred. No.: 1.54e-47 Length: 230
Score: 599.50 Matches: 118
Percent Similarity: 93.13% Conservative: 4
Best Local Similarity: 90.08% Mismatches: 6
Query Match: 4.98% Indels: 3
DB: 4 Gaps: 1

US-10-664-859-1 (1-6909) x US-09-270-767-44999 (1-230)

QY 2382 TCATTTTACATAGTTAAATTTGAAGAATTTTCAACGACAGTACCACGGAATAAAGGA 2441
Db 103 SerPheTyr***SerThrValLysLeuMetSer***AsnGly-----LysLysGly 119
QY 2442 TCGTCCTTGACATGATAATGACGAATGAGCATGGAAGCTGCAATCAGTTGAATCCC 2501
Db 120 SerSerLeuThrMetAsnAsnAspGluMetSerMetGluGlyCysAsnGlnLeuAsnPro 139
QY 2502 GATTTTATCAATGAATCTTTAATAATCTTGAATTTTGAGCATATTAAGCGGAGTA 2561
Db 140 AspPheIleAsnGlnSerLeuAsnAsnProAlaIleSerSerIleLeuValSerGlyVal 159
QY 2562 GGACCAATATCCCGAATCGAGTTGAGCGGGACGGGAATTTATGACTGCCAACGCC 2621

Db 160 GlyProIleProGlyIleGlyValGlyAlaGlyThrGlyAsnLeuLeuThrAlaAsnAla 179
QY 2622 AATGAATCTCCTCGGGTAGCAGTAATTGTTGGATTACATGCAACAGCAAAATCACATA 2681
Db 180 AsnGlyIleSerSerGlySerSerAsnCysLeuAspTyrMetGlnGlnGlnAsnHisIle 199
QY 2682 TTGCTGTTTCAACTCAGCTGGCCCAACAAAGGGCCGAATCAGTTTAAAGCGTCAATT 2741
Db 200 PheValPheSerThrGlnLeuAlaAsnLysGlyAlaGlnSerValLeuSerGlyGlnPhe 219
QY 2742 CAACTATTATTGCGTATCACTGACTCAGCCT 2774
Db 220 GlnThrIleIleAlaTyrHisCysThrGlnPro 230

RESULT 7

US-09-270-767-36764
; Sequence 36764, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 36764
; LENGTH: 140
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-36764

Alignment Scores:
Pred. No.: 3.8e-47 Length: 140
Score: 594.00 Matches: 123
Percent Similarity: 88.57% Conservative: 1
Best Local Similarity: 87.86% Mismatches: 16
Query Match: 4.93% Indels: 0
DB: 4 Gaps: 0

US-10-664-859-1 (1-6909) x US-09-270-767-36764 (1-140)

QY 6391 CACCAGACACATGATTAACAATGCCCGCAATTATTCACCAAAATCCACGTTCTTTGTC 6450
Db 1 HisGlnSerAsnMetIleThrMetProSerAsnLeuSerProAsnProThrPhePheVal 20
QY 6451 AACAAATAACTTCTTAATTTTGGCCGCTCGTCATGTATGTTTACTAGTCTCCAAT 6510
Db 21 AsnLys***ThrSerLysPheLeuProSerLeuProCysIleValTyr***SerProAsn 40
QY 6511 TAAGACATGCATCTTAATTAAGATTTTGAAGCTTATTTACTTACGTTTGAAC 6570
Db 41 ***AspMetHisLeu***IleArgPhePheGluAlaTyrLeuLeuArgCysPheTyrAsn 60
QY 6571 GGAGAAATTAACCTTTTGATATGCAATGATAAGTTGGAACAACATAATTCAATTGTC 6630
Db 61 GlyGlnAsnLysLeuLeuAspMetGlnMetIleThrLeuGluThrThr***PheIleCys 80
QY 6631 AACTTTTGAAGTCACGTCGAAGTTAAATGTAGAATCTGTATTTTAAACATAATAGTCAT 6690
Db 81 AsnPhe***LysSerArgArgSer***Met***AsnLeuTyrPheAsnIleIleGlyHis 100
QY 6691 CTGTAATAATAATTAACATCGAAATTTTAGTTATCAGCAGCTATTTCTGTTATTAATT 6750
Db 101 LeuLeuLys***LeuAsnIleGluIleLeuValIleSerSerTyrTyrLeuLeuLeuPhe 120
QY 6751 AATATGTGCGCTGCTCTCTGTGTTAATGAATTAATAATATATATAATGTAAC 6810
Db 121 AsnMetCysAlaAlaLeuSerValLeuAsnGluIleLysIleTyrIle***Met***Asn 140

```
RESULT 8
US-09-270-767-51981
; Sequence 51981, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 51981
; LENGTH: 140
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-51981

Alignment Scores:
Pred. No.:      3.8e-47      Length:      140
Score:          594.00      Matches:      123
Percent Similarity: 88.57%      Conservative: 1
Best Local Similarity: 87.86%      Mismatches: 16
Query Match:    4.93%      Indels:      0
DB:             4          Gaps:      0

US-10-664-859-1 (1-6909) x US-09-270-767-51981 (1-140)

QY      6391 CACCAGAGCAACATGATAACAATGCGCCGGAATTATCACCMAATCCAACGTTCTTGTG 6450
      1 HisGlnSerAsnMetIleThrMetProSerAsnLeuSerProAsnProThrPhePheVal 20
QY      6451 AACAAATAAAGCTTCTAAATTTTGGCCGCGCTCGTCATGTATTGTTTACTAGTCTCCAAT 6510
      21 AsnLys***ThrSerLysPheLeuProSerLeuProCysIleValTyr***SerProAsn 40
QY      6511 TAAGACATGCATCTCTAAATAAGATTTTGAAGCTTATTACTTAGTGTTTACCAAC 6570
      41 ***AspMetHisLeu***IleArgPhePheGluAlaTyrLeuLeuArgCysPheTyrAsn 60
QY      6571 GGAGAAAATAACTTTTGGATATGCAAAATGATTAACGTTGGAACAACATAATTTCATTGC 6630
      61 GlyGlnAsnLysLeuLeuAspMetGlnMetIleThrLeuGluThrThr***PheIleCys 80
QY      6631 AACTTTGAAGTCACGTCGAAGTTAAATGAGAATCTGTATTTTTAACAATAAGTCAT 6690
      81 AsnPhe***LysSerArgArgSer***Met***AsnLeuTyrPheAsnIleIleGlyHis 100
QY      6691 CTGTAAAATAATTAACATCGAAATTTTAGTTATCAGCAGCTATTTTCTGTATTATT 6750
      101 LeuLeuLys***LeuAsnIleGluIleLeuValIleSerSerTyrTyrLeuLeuLeuPhe 120
QY      6751 AATATGTGCGCTGCTCTCTCTGTGTTAAATGAATTAATATATATATATAATGTAAC 6810
      121 AsnMetCysAlaAlaLeuSerValLeuAsnGluIleLysIleTyrIle***Met***Asn 140

RESULT 9
US-09-270-767-43686
; Sequence 43686, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 43686
; LENGTH: 65
; TYPE: PRT
```

```
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-43686

Alignment Scores:
Pred. No.:      9.59e-21      Length:      65
Score:          312.00      Matches:      63
Percent Similarity: 96.92%      Conservative: 0
Best Local Similarity: 96.92%      Mismatches: 2
Query Match:    2.58%      Indels:      1
DB:             4          Gaps:      0

US-10-664-859-1 (1-6909) x US-09-270-767-43686 (1-65)

QY      4766 ATCGTACATCTTATTTGGCTCTATAGACTATTTACTGGTGACCGCTGACCGCTGTGGCC 4707
      1 IleValHisLeuTyrLeuProLeuLeuAspTyrLeuLeuValThrAlaAspProValGly 20
QY      4706 TTGGTCGCCCTACTTTTCATGTTATCACCTGTAGAATATTAATAACATATTTTAAT 4647
      21 LeuValAlaTyrPheSerCysTyrHisLeu***LysTyrTyrAsnLysHisIlePheAsn 40
QY      4646 TTCTTATTAACAATAGTAGTATTTTCAATTAAC-GGATATGAATTTTCGTGTTAATTTTA 4588
      41 PheLeuLeuAsnAsnSerAspPheGlnLeuThrGlyTyrGluPheSerCysLeuIleLeu 60
QY      4587 ATAGTAGATTATCTC 4573
      61 IleValAspTyrLeu 65

RESULT 10
US-09-949-016-7088
; Sequence 7088, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for windows Version 4.0
; SEQ ID NO 7088
; LENGTH: 1179
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-7088

Alignment Scores:
Pred. No.:      4.52e-13      Length:      1179
Score:          238.50      Matches:      250
Percent Similarity: 30.38%      Conservative: 137
Best Local Similarity: 19.62%      Mismatches: 406
Query Match:    1.98%      Indels:      481
DB:             4          Gaps:      64

US-10-664-859-1 (1-6909) x US-09-949-016-7088 (1-1179)

QY      2814 CCTTAAAGATTAAACAAGTTACAGCGGCAATTCGCTCGGTATGCCATGATAGGCATG 2873
      156 ProSerGlnGlnSerThrThrGlnGlnHisProGlnSerGlnProProTyr----- 172
QY      2874 GGGCAGGTTGACTACTCTCTCTAATCTGTAGCCAAATAACACAACAGCAGCCACAT 2933
      173 -----SerGlnProGlnAlaGlnSerProTyrGlnGlnGlnGlnProGln 187
```

```
QY 2934 ACAAGACCGTAGCCATTG-----AAACCCCAATTCAATCAA 2972
Db 188 GlnProAlaProSerThrLeuSerGlnGlnAlaIaIaTyrrProGlnProGlnSerGlnGln 207
QY 2973 CATGAAAACAGCAAAACGTAAGTACTGTAAGCGCCGCTAGCAACTCTTTGTCGACCACTCT 3032
Db 208 SerGlnGlnThrAlaIaTyrrSerGlnGlnArgPheProProGlnGlnLeuSerGln--- 226
QY 3033 GATCCTTAGGCAACGAACTGAATTGATGTCGTGGAAAGCGGATCCTCAACACACAGT 3092
Db 227 AspSerPheGlySerGlnAlaSer-----SerAlaProSerMetThrSerSer 242
QY 3093 AGGTCTGACAAAACTCAGAAATCATGTAGACAGTATCAGTACATCCAGCGAGTCACAG 3152
Db 243 LysGlyGlyGlnGlnIuAspMetAsn-----LeuSerLeuGlnSerArgProSer 258
QY 3153 GCAATAAAGATACTGGAAGCAGCTGGCGTGTGATTGGGACAGGTCACAAAGAAC--- 3209
Db 259 SerLeuProAspLeuSerGlySerIleAspAspLeuProMetGlyThrGlyGlyAlaLeu 278
QY 3210 GATCCTGGCCTGACAACTGAAAACAACATTTGTATCACTGCAAGGAGTTAAGTTCCAGAC 3269
Db 279 SerProGlyValSerThr---SerGlyIleSerSerSerGlnGly----- 292
QY 3270 GAAAAACCTTACACCAACAG-----CGGCAACATCGGGAAGAACAGTTGGCAAAA 3320
Db 293 GluGlnSerAsnProIaGlnSerProPheSerProHisThrSerProHisLeuProGly 312
QY 3321 ATAAAAAAATGAATCAATTTCTTTTCTGMAAATGAGAATTCAAGTAGAGATAATGTA 3380
Db 313 IleArg-----GlyProSerProSerProValGlySer----- 323
QY 3381 AGCTCAGATAAACAAAATTCCAGAGATTTAATGATGGGATGTCGGGTGGCGAGGC 3440
Db 324 -----ProIaSerValaIaGlnSerArgSerGlyPro----- 334
QY 3441 GGATCTATTATAATCCGACGATGCGACAACTGCATATGCCAGGTAACGCCAAATCGGAG 3500
Db 335 -----LeuSerProAlaIaIa-----ValProGlyAsn-----Gln 344
QY 3501 CTCTTATCGGCGACAACTTCAGGACTTTCGGAAGATGTAATGCATCCA----- 3548
Db 345 MetProProArgProProSerGlyGlnSerAspSerIleMetHisProSerMetAsnGln 364
QY 3549 GGGGATGTATATCAGATATGGGTGCGCTAATAGATGTATAATAAT----- 3596
Db 365 SerSerIleAlaGlnAspArgGlyTyrrMet-----GlnArgAsnProGlnMetPro 381
QY 3597 CAAAAAACAGTGTGCAATGTGATCTGAGTAGTGTTGTCACTGGAACAACCTGCAGCT 3656
Db 382 GlnTyrrSerSerProGlnProGlySerAlaLeu----- 392
QY 3657 GGAGTAATGTCAATATGCATTGCTCAAGCTCCGGCGCCCGAATGGCAATATGATGGGA 3716
Db 393 -----SerProArgGlnProSerGlyGlyGlnIleHis 403
QY 3717 AGCTTACGATATGTAGCCTCGTTTGGCAACACAGCTGCAACGTCATCGGAACGGCC 3776
Db 404 Thr-----GlyMetGlySerTyrrGlnGlnAsnSerMetGly----- 415
QY 3777 CCAGATATGTCTAAGGAATTAAATCAAGATAGCCGAACCCATTCAATCAACAGGGGA 3836
Db 416 -----SerTyrrGlyProGlnGlyGly 422
QY 3837 GTTGCTCAAAATGAGTGTGGAAGATTCAACATCAATTTTTCGAAGAAGCGCTCAAGGG 3896
Db 423 -----GlnTyrrGlyProGlnGly 428
QY 3897 GGCAGAGCCAGACAAGTCACTGGAAGTGTAGTACACAACAGCAAAACCCCTTCT----- 3950
Db 429 GlyTyrrProArgGlnProAsnTyrrAsnAlaLeuProAsnAlaAsnTyrrProSerAlaGly 448
```

```
QY 3951 -----GGATCTGGTGAAACTCGTTAAACAACAGGTGCGGA 3986
Db 449 MetAlaGlyIleAsnProMetGlyAlaGlyGlyGln----- 461
QY 3987 CCCCTGCAAGGTCCACT-----CCTCCTTAACCACTCCATCCAGAGATCTGGGTCACTA 4040
Db 462 ---MetHisGlyGlnProGlyIleProProTyrrGlyThrLeuPro---ProGlyArgMet 479
QY 4041 CCAATAGCCCACTCAATCGCCCAATCCCTCGAGTCCAACAATCTATCTCTCCG----- 4094
Db 480 SerHisAlaSerMetGlyAsnArgProTyrrGlyProAsnMetAlaAsnMetProProGln 499
QY 4095 -----TCACCGCGGACAACCGCA 4112
Db 500 ValGlySerGlyMetCysProProProGlyGlyMetAsnArgLysThrGlnGlnThrAla 519
QY 4113 GCAGTCATGGGATTGCCGACCAACTCT-----CCTAGCATG 4148
Db 520 ValAlaMetHisValAlaAlaAsnSerIleGlnAsnArgProProGlyTyrrProAsnMet 539
QY 4149 GAT-----GGAAACAGGA-----TCATATATCT 4169
Db 540 AsnGlnGlyGlyMetMetGlyThrGlyProProTyrrGlyGlnGlyIleAsnSerMetAla 559
QY 4170 GGATCTGTCCGCAAGCTAATACTTCGACGGTTCAAGCAGGACAGCACAACAAGTCTCTCA 4229
Db 560 GlyMetIleAsnProGlnGlyProProTyrrSerMetGlyGlyThr-----Met 575
QY 4230 GCAACAAGAACTGTTTTCAGGCAGACACCCCA----- 4262
Db 576 AlaAsnAsnSerAlaGlyMetAlaAlaAsnSerIleGlnAsnArgProGluMetMetGlyLeuGlyAspValLys 595
QY 4263 ---TCGCCGTCAAATCAAATCGTAGTAGAAATACCGGATCGTCAAGCGTTCTTACCGCAT 4319
Db 596 LeuThrProAlaThrLysMetAsnAsnLysAlaAspGlyThrProLysThrGlnSerLys 615
QY 4320 AACTTAAGCAGCAACCCAACTACCCCTTATCTCATCTATCCCAAAGGAATTGAGTCT 4379
Db 616 SerLysLysSerSerSerSerThrThrThrAsnGlnLysIleThrLysLeuTyrrGlnLeu 635
QY 4380 TTCGCTAGTCTCTCTGTGTATG-----TTATATTTGTTAATTTTTTAAAGAC 4430
Db 636 GlyGlyGlyProGluArgLysMetTyrValAspArgTyrrLeu---AlaPheThrGlnGlu 654
QY 4431 AAA-----TCAATATGAAATTCGCTTAATAAAG----- 4460
Db 655 LysAlaMetGlyMetThrAsnLeuProAlaValaGlyArgLysProLeuAspLeuTyrrArg 674
QY 4461 TTATATATTACA-----TAACTCGGAATTTGATGAAAAAATCAGGAATAGAAAAAAT 4514
Db 675 LeuTyrrValSerValLysGluIleGlyGlyLeuThrGlnValAsnLysAsnLysLysTrp 694
QY 4515 AAATTATTTTCCGGAACCGCCCATCTCATTTCTGAATCCAATTTCTGAGTGATTGTAGA 4574
Db 695 Arg----- 695
QY 4575 GATAATCTACTAATTAAATTAAACAGAAATTCATATCCGTTAATTGAAATCACTATT 4634
Db 696 ---GluLeuAlaThrAsnLeuAsnValGlyThrSer-SerSerAlaAlaSerSerLeu-- 713
QY 4635 GTTATATAGAAATTAATAATATGTATTATAAT----- 4669
Db 714 ---LysLysGlnTyrrIleGlnCysLeu-TyrrAlaPheGluCysLysIleGluArgGlyG 732
QY 4670 -----ATTTCACAGGTGATACATGAAGT-----A 4697
Db 732 IuAspProProProAspIlePheAlaAlaAspSerLysLysSerGlnProLysIleG 752
QY 4698 GCGGACCAAGCCA-----CAGGTCAGCGGTCAACCACTAATAGTC 4739
Db 752 IuProProSerProAlaGlySerGlySerMetGlnGlyProGlnIuProGlnSerThrs 772
QY 4740 TAATAGAGGCAAAATAAGATGTAGATTGCTGCATCCAGTCCGTGTTTAAACCCGCAT- 4798
```



```

; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949, 016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241, 755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237, 768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231, 498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7421
;
; LENGTH: 1076
;
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-7421

```

```

Db      495 ArgProThrSerArgLeuAsnArgLeuProGluLeuGluLeuGluAlaIleAspAsnGln 514
Qy      4278 AATCGTAGTAGAAATACCGGA-----TCGTCAAGCGTCTTACGCAT 4319
Db      515 PheGlyGlnProGlyThrGlyAspGlnIleProThrAsnAsnThrValThrAlaIle 534
Qy      4320 AACTTAAGCAGCAACCA-----AGTACCCCTTATCTCATCTA-----TCC 4361
Db      535 AsnGlnSerLysSerGluAspGlnCysIleSerSerGlnLeuAspGluLeuLeuCysPro 554
Qy      4362 CCAAGGAATTGAGTCTTCCGTCACTCTGCTGCTGATAGTTATTTGTTAATTTT 4421
Db      555 ProThrThrValGluGlyArgAsnAspGluLysAlaLeuLeuGluGlnLeuValSerPhe 574
Qy      4422 TTT-----AAGACAATCAATATGAAATGCGTTAATAATGTTATATATTACATAA 4475
Db      575 LeuSerGlyLysAspGluThrGluLeuAlaGluLeuAspArgAlaLeuGlyIleAspLys 594
Qy      4476 CTC-----GGAATTTGATAGAAAAATCAGAAATAGA----- 4508
Db      595 LeuValGlnGlyGlyGlyLeuAspValLeuSerGluArgPheProGlnAlaThr 614
Qy      4509 -----AAAAATAAATTATTTC-----GGA 4529
Db      615 ProProLeuIleMetGluGluArgProAsnLeuTyrSerGlnProTyrSerSerProSer 634
Qy      4530 CCGCCCATCATTTCTTGAATCCAAATTTCTGAGTGATGTTAGATTAATCTACTATTA 4589
Db      635 ProThrAlaAsnLeuProSerProPheGlnGlyMetValArgGlnLysProSerLeuGly 654
Qy      4590 AAATTAACACGAAAAATTCATATCCGTTAATTGAAATCACTATTGTTAATAAGAAATT 4649
Db      655 ThrMetProValGlnVal--ThrPro----- 662
Qy      4650 AAAAATATGTTTATTATATATTCTACAGTGATACATGAAGAAGTAG----- 4699
Db      663 -----ProArgGlyAlaPheSerProGlyMetGlyMetGlnProArgGlnThrLeuA 680
Qy      4700 --CGACCA-----AGCCCA-----CAGGCTC 4718
Db      680 snArgProProAlaAlaProAsnGlnLeuArgLeuGlnLeuGlnArgLeuGlnGlyG 700
Qy      4719 AGCGGTACACAGTAATAGTCTAATAGAGCAATAAAGATGTA-----CGATTGG 4769
Db      700 InGln-----GlnLeuIleHisGlnAsnArgGlnAlaIleLeuAsnGlnPhea 716
Qy      4770 CTGCATCCAGTCT--GTTTAAACCCGCATCCACATATGCAAGC-----AATTCAA 4820
Db      716 IaAlaThrAlaProValGlyIleAsnMetArgSerGlyMetGlnGlnIleThrProG 736
Qy      4821 ATTCAGCATTAACGCTATAAATGGGCTCTACCAATATACAGATGAGGTAATATTT 4880
Db      736 InProProLeuAsnAlaGlnMetLeuAlaGlnArgGlnArgGluLeu----- 751
Qy      4881 AAATATTTTATTAAAGTTTGTGTTAATTATCTTTTTCAGCGTCAAGCATCAGC 4940
Db      752 -----TyrSerGlnGlnHisArgG 758
Qy      4941 GCAAGGTGATCCGTACAATTAGTCGGCGCTCCGATATATTCCGTAATCCCAATAG 5000
Db      758 In-ArgGlnLeu----- 761
Qy      5001 TGGCAATCGCCGCCAACCAACAGATGACCAAAACTTCGATCCAATCTTCTTTGGC 5060
Db      762 -----IleGlnGlnIleArgAlaMetLeuMetArgGlnGlnSerPheGly 776
Qy      5061 ACAATGTCCCAACAACAATAACAAGTTGCGTGTCCAGCATGGTAGTCCAGCCGAAGTGG 5120
Db      777 AsnAsnLeuProProSerSerGlyLeuProValGln-MetGly----- 790
Qy      5121 TGGTATGACGATGATGGGGGTCCGGGACCGTCCGACATCATATATTGACATGAATAAT 5180
Db      790 ----- 790

```

```

Qy      5181 TTCGGAGCTAGATGATCAGGAATAGATACCATAAATCAAAATAACTGTCAATTCAATGAA 5240
Db      790 ----- 790
Qy      5241 TGTGTAATGAATCAATGGGTCCCGAATGCTGAATCCTAAATGTGCGTAGCAGCGCG 5300
Db      791 -----AsnProArgLeu----- 794
Qy      5301 TCCAATGACCGCT-----GGCTTAATCCTAATTC 5333
Db      795 -ProGlnGlyAlaProGlnGlnPheProTyrProProAsnTyrGlyThrAsnProGlyTh 814
Qy      5334 CCCCAATGGTGATTAGAGAGAAATTCATAGGGTCTGGCTGTGCTCAGCAAACTCTTC 5393
Db      814 rProProAlaSerThrSerProPheSerGlnLeuAlaAlaAsnProGluAlaSerLeuAl 834
Qy      5394 AAACCTTCAAGGGGTGTTCCACCTGCTGTCAGAAATGATGGTCCAATGCAATTT 5453
Db      834 aaAsnArgAsnSerMetValSerArgGly-----MetThrGlyAsnIleGlyGlyGlnPh 852
Qy      5454 TGGTTCGAATTTCAATCCGAATATTCAAGTAAGGCGAGTACCCCAACACCATACATA 5513
Db      852 eGlyThrGlyIleAsnProGlnMetGln-----GlnAsnValPheGlnTyr 867
Qy      5514 C-----ATCGCAGTAAGGCGACAGAACCGCAACAACAATAACAACAATGG 5558
Db      867 rProGlyAlaGlyMetValProGlnGlyGluAlaAsnPheAlaProSerLeuSerProGl 887
Qy      5559 AGCTAATAATGTGCGAATG-----CCACCTAGTCTGGAATTTTTCAGAGG----- 5604
Db      887 ySerSerMetValProMetProIleProProGlnSerSerLeuLeuGlnGlnThrPr 907
Qy      5605 -----TAGCTAACCT-----CAATGGGTGCTGTAGGCAA 5636
Db      907 oProAlaSerGlyTyrGlnSerProAspMetLysAlaTrpGlnGlnGlyAlaIleGlyAs 927
Qy      5637 TGGTCGCCAATATGCCCACCATCAGCCAGCGACTACTCTGGAATGCCAGATGG-- 5694
Db      927 nAsnAsnValPheSerGlnAlaValGlnAsnGlnProThrProAlaGlnProGlyValTy 947
Qy      5695 -----ATGGCGGACAGCAGCGGAGGTATGCTAATGAA 5729
Db      947 rAsnAsnMetSerIleThrValSerMetAlaGlyIleAsnThrAsnValGlnAsnMetAs 967
Qy      5730 TTCTCCGAGAGACACACACAGACAAGATCACAAACAATCCTGGGGCAAGC----- 5781
Db      967 nProMetMetAlaGlnMetGlnMetSerSerLeuGlnMetProGlyMetAsnThrValCy 987
Qy      5782 -----AATGTAATTAATTCTTTCAGAAATG 5807
Db      987 sProGluGlnIleAsnAspProAlaLeuArgHisThrGlyLeu-----TyrCy 1003
Qy      5808 CAATCAATGTCTATTGTTGAC-----GAAGAGGTGATTAACCGGCATGACGG 5858
Db      1003 sAsnGlnLeuSerSerThrAspLeuLeuLysThrGluAla-----AspG 1018
Qy      5859 ATCAATGAATATTGTCACACCATATG-----ATAAG 5891
Db      1018 yThrGlnGlnValGlnGlnValGlnValPheAlaAspValGlnCysThrValAsnLeuVa 1038
Qy      5892 GGGCATGCGTCCACATGCGCGCAATGTAATGGGTGCGCGGATGCCACCGTTAA 5951
Db      1038 lGlyGlyAspProTyrLeuAsnGlnProGlyProLeuGlyThrGlnLysProThrSerG 1058
Qy      5952 CAGGCAAAATTCAGTTGCACAG 5973
Db      1058 yProGlnThrProGlnAlaGln 1065

```

RESULT 13
 US-09-125-635-4
 ; Sequence 4, Application US/09125635
 ; Patent No. 6562589


```

; GENERAL INFORMATION:
; APPLICANT: THE UNITED STATES OF AMERICA represented by THE SE
; TITLE OF INVENTION: AIB1, A novel steroid receptor co-activator
; FILE REFERENCE: 49944
; CURRENT APPLICATION NUMBER: US/09/125,635
; PRIOR FILING DATE: 1998-08-21
; PRIOR APPLICATION NUMBER: 60/049,728
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1420
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-125-635-4

```

```

Alignment Scores:
Pred. No.: 1.81e-11 Length: 1420
Score: 222.00 Matches: 267
Percent Similarity: 30.04% Conservative: 150
Best Local Similarity: 19.24% Mismatches: 496
Query Match: 1.84% Indels: 475
DB: 4 Gaps: 65
US-10-664-859-1 (1-6909) x US-09-125-635-4 (1-1420)

```

```

QY 3120 GTAGACAGTATCATCATCCAGCGAGTCCAGGCAATAAAGATCTGGAAGCAGCTGGC 3179
Db 95 ValGlnLysAlaAspValSerSerThrGlyGlnGlyValIleAspLysAspSerLeuGly 114
QY 3180 GTTGATTGGGACAGTCACAAAAGAGCCGATCTGGCTGACAACTGMAAACAACATT 3239
Db 115 ProLeuLeuGlnAlaLeuAspGlyPheLeuPheValValAsnArgAspGlyAsnIle 134
QY 3240 GTATCACTGCAAGAGTTAAGTTCCAGAGAAACCTTACACCAACAAGCGGCAACAT 3299
Db 135 ValPheVal-----SerGluAsnValThr--GlnTyrLeuGlnTyr 147
QY 3300 CGGAGAAGACAGTTGGCAAAAATAAAAAATGAATCAATTTCTTTTCTGAAATGAG 3359
Db 148 LysGlnGlnAspLeuValAsnThrSerValTyrAsn--IleLeuHisGlnGluAspArg 166
QY 3360 AATTCACTAGAGCTAATGTAAGCTCACAGATAACAATAATCCAGAGATTAAATGATG 3419
Db 167 LysAspPhe-----LeuLysAsnLeuProLysSerThrValAsn 179
QY 3420 GGGATGTCGGGTGGCGGAGCGGATCTATTATTAATCCGACGATCGGACAACTGATATG 3479
Db 180 GlyValSerTyr-----ThrAsnGlnThrGlnArgGlnLysSerHis 193
QY 3480 CCAGGTAACGCCAAATCGGAGCTTATCGGCAACAAGTTCAGAGCTTTGGAGAAGATGTA 3539
Db 194 ThrPheAsnCysArg-----MetLeuMetLysThr----- 203
QY 3540 ATGCATCCAGGGATGTTATATCAGATATGGTGCCGTAATAGATGTAATAATCAATCAA 3599
Db 204 -----ProHisAspIleLeuGlnAspIleAsnAlaSerProGluMetArgGlnArgTyr 221
QY 3600 AAAACCAAGTGTGCAATGTGATCTGAGTAGGTGTGTCACTGGAACAACCTGCACTGGA 3659
Db 222 GluThr----- 223
QY 3660 GTAATGTCAATATGATGTCTCAAGCTCCGGCGCCCGAATGGAATATGATGGGAAGC 3719
Db 224 -----MetGlnCysPheAlaLeuSerGlnProArgAla---MetMetGluGlu 238
QY 3720 TCTACGATATGCTAGCCTCGTTTGGCAACAAGCTGCAACGTC----- 3764
Db 239 GlyGlnAspLeu-----GlnSerCysMetIleCysValAlaIaArgArg 252
QY 3765 -----ATCGGAACGGCGCCAGAT 3782
Db 253 IleThrThrGlyGlnArgThrPheProSerAsnProGluSerPheIleThrArgHisAsp 272

```

```

QY 3783 ATGCTAAGCAAGTTTAAATCAAGATAGCCGAACC-----CATTCACATCAAGGGGA 3836
Db 273 LeuSerGlyLysValValAsnIleAspThrAsnSerLeuArgSerSerMetArgProGly 292
QY 3837 GTTGCTCAATG-----GAG 3851
Db 293 PheGluAspIleIleArgArgCysIleGlnArgPhePheSerLeuAsnAspGlyGlnSer 312
QY 3852 TGGTCGAAGATTCACATCAATTTTTCGAGAAGACGCTCAAGGGGGGCAAGCCAGACAA 3911
Db 313 TrpSerGlnLysArgHis--TyrGlnGlnAlaTyrLeuAsnGlyHisAlaGluThrPro 331
QY 3912 GTC-----ACTGGAAGTGTAGTACCAACAACAGCAAAACCCCTTCTGGA 3953
Db 332 ValTyrArgPheSerLeuAlaAspGlyThrIleVal---ThrAlaGlnThrLysSerLys 350
QY 3954 TCTGTGGAACCTCGTTAACAACACAG----- 3980
Db 351 LeuPheArgAsnProValThrAsnAspArgHisGlyPheValSerThrHisPheLeuGln 370
QY 3981 -----GTGCGACCCCTG-----CAAGTCCACCTCTCTCT 4010
Db 371 ArgGlnGlnAsnGlyTyrArgProAsnProAsnProValGlyGlnGlyIleArgProPro 390
QY 4011 TACCACTCCATCCAGAGATCTGCGTCAGTACCAATAGCCACTCAATCGCCCAAT----- 4064
Db 391 MetAlaGlyCys-----AsnSerSerValGlyGlyMetSerMetSerProAsnGlnGly 408
QY 4065 -----CCCTCGAGTCCAAACAATGTATCTCTCCCGTCACCGCGGCAACCGCAGCA 4115
Db 409 LeuGlnMetProSerSer--ArgAlaTyrGlyLeuAlaAspProSerThrThrGlyGln 427
QY 4116 GTCATGGGATGCGC-----ACCAACTCTCCTAGC 4145
Db 428 MetSerGlyAlaArgTyrGlyGlySerSerAsnIleAlaSerLeuThrProGlyProGly 447
QY 4146 ATGATGGAACAGATCATATTCTGATCTGTCGCGCAAGCTAATACTTCGACGGTTACG 4205
Db 448 MetGlnSerProSerSerTyrGlnAsnAsnAsnTyrGlyLeuAsnMetSerSerProPro 467
QY 4206 GCAGGCACAACAACAGTCTCTCAGCAACAAGAACTGTTTTCAGGAGACACCCCATCG 4265
Db 468 HisGlySerProGlyLeu-----Ala 474
QY 4266 CCGTCAATCAAAAT-----CGTAGAATAATACCGGATCGTCAAGCTTCTTACG 4316
Db 475 ProAsnGlnGlnAsnIleMetIleSerProArgAsnArgGlySerProLysIleAlaSer 494
QY 4317 CATTAAGTAAAC-----AGCAACCAAGTACCCCTTATCTCATCTATCCCAAGGAA 4370
Db 495 HisGlnPheSerProValAlaGlyValHisSerProMetAlaSerSerGlyAsnThrGly 514
QY 4371 TTGAGTCTTTCGCTCAGTCTCTGCTGATGTTATATTGTTAAAT----- 4418
Db 515 AsnHisSerPheSerSerSerSerLeuSerAlaLeuGlnAlaIleSerGlnGlyValGly 534
QY 4419 -----TTTTTTAAAGACAAATCAAAATATG 4442
Db 535 ThrSerLeuLeuSerThrLeuSerSerProGlyProLysLeuAspAsnSerProAsnMet 554
QY 4443 AATTGCGTTAATAATAGTATATATTAATCACTCGGAATTTGATAGAAAAAATCAGG 4502
Db 555 AsnIleThrGlnPro-----SerLysValSer 563
QY 4503 AATAGAAAAATAAATTATTTTCCGAGCGCCCATCATTTCTTG-----AATCCA 4553
Db 564 AsnGlnAspSerLys-----SerProLeuGlyPheTyrCysAspGlnAsnPro 579
QY 4554 ATTTCTGAGTGAATTGTT-----AGAGATAATCTACTATTAAAA----- 4592
Db 580 ValGlnSerSerMetCysGlnSerAsnSerArgAspHisLeuSerAspLysGlnSerLys 599

```

```
OY 4593 -----TTAACACGAAATTCTATC 4613
Db 600 GluSerSerValGluGlyAlaGluAsnGlnArgGlyProLeuGluSerLysGlyHisLys 619
OY 4614 CGTTAATTGAAATCACTATTGTTTAATAGAAATTAAATAATATGTTATTATATATT 4673
Db 620 Lys-----Leu 621
OY 4674 CTACAGGTGATTAACATGAAAAGTAGCGACCAAGCCACAGGTCAGCGTCACCAGTAA 4733
Db 622 LeuGlnLeuLeuThrCysSerSerAspArgGlyHis-SerSerLeuThrAsnSer-- 640
OY 4734 ATAGTCTAATAGAGCAATAAAGATGTACGATTGCTGCATCCAGTCCGTGTTTAACC 4793
Db 641 -----Pr 641
OY 4794 CGCATCCACATATGCAAGCAATTCAATTACGATTAAACGCCATTAATGGCTCTA 4853
Db 641 oleuAspSerSerCysLysGluSerSerValSerValThrSerPro---SerGlyValSe 660
OY 4854 CCAATATACAGATGAGGT-----AAATATTTAATATTTATTATTAA 4895
Db 660 rSerSerThrSerGlyGlyValSerSerThrSerAsnMetHisGlySerLeuLeuGlnG 680
OY 4896 CGTTTGTGTTAATTATCTCTTTTTCAGCGT----- 4929
Db 680 uLysHisArgIleLeuHisLysLeuLeuGlnAsnGlyAsnSerProAlaGluValAlaLy 700
OY 4930 -----CAAGCATCAGCGCAAGGTGATCCGTACAAATTGATCGGCGCTCCGATAA 4979
Db 700 sIleThrAlaGluAlaThrGlyLysAspThrSerSerIleThrSerCysGlyAspGlyAs 720
OY 4980 TATTCGCTAAATCCCAATAGTGGCAATCGGCGCCACCAACAAGATGACCCAAAACCT 5039
Db 720 nValValLysGlnGlnLeuSerProLysLysGluAsnAsnAlaLeuLeuArgTy 740
OY 5040 C-----GATCCAATCTCTTTGGCA----- 5061
Db 740 rLeuLeuAspArgAspAspProSerAspAlaLeuSerLysGluLeuGlnProGlnValG 760
OY 5062 -----CAATGTCCCAACAACATAACAAGTTGCGTCCGACATGGTAGTCC 5108
Db 760 uGlyValAspAsnLysMetSerGlnCysThrSerThrIleProSerSerSerGlnG 780
OY 5109 AGCCGGAAGTGGTGTATGACGATGATGGGGGTCGGGACCGTCGACATCAATATTGA 5168
Db 780 uLysAspProLysIleLysThrGluThrSerGlnGlnLysSerGlyAspLeuAspAsnLe 800
OY 5169 GCATGAATTAATTTCCGGAAGTAGATGATGACGAATA-----GATACCAATAATCAAAA 5222
Db 800 uAspAlaIleLeuGlyAspLeuThrSerSerAspPheTyraAsnAsnSerIleSerSerAs 820
OY 5223 TAACTGTCAATTCAATGAATGTCGTAATGAATCAATGGTCCCGAATGCTGAATCCTAA 5282
Db 820 nGlySerHis-----LeuGlyThrLy 827
OY 5283 AATGTGCTAGCAGCGGTCCTCAATGCA----- 5310
Db 827 sGlnGlnValPheGlnGlnGlyThrAsnSerLeuGlyLeuLysSerSerGlnSerValGlnSe 847
OY 5311 -----CCGCCT-----GGCTTTAATCCTAATTCGCCCAATGGTGATTAGAGA 5354
Db 847 rIleArgProProTyraAsnArgAlaValSerLeuAspSerPro----- 861
OY 5355 GAATTCATAGGCTCTGGCTGTGGCTCAGCAAACTCTTCAAACTTTCAAGGGGTTGTTCC 5414
Db 862 -ValSerValGlySerSerProProValLysAsnIleSerAlaPhePro-----MetIle 879
OY 5415 ACCTGTGCGCAGATGATGGTGCATGCCA-----GTCAATTTTGG 5456
Db 879 uProLysGlnProMetLeuGlyGlyAsnProArgMetMetAspSerGlnGluAsnTyrgI 899
OY 5457 TTGCAATTTC-----AATCCGAATATTCAAGTAAAGCGAGTACCCCAACACCAT 5507
```

```
Db 899 ySerSerMetGlyGlyProAsnArgAsnValThrVal---ThrGlnThrProSerSerG 918
OY 5508 ACATATC---ATGCCA-----GTAAGGCGCACAGCGCCAAACAACA 5546
Db 918 yAspTrpGlyLeuProAsnSerLysAlaGlyArgMetGluProMetAsnSerAsnSerme 938
OY 5547 TAACACAATGGAGCTAAT---AATGTCCGAATGCCACCTAGTCTGAATTTTGCAGAG 5603
Db 938 tGlyArgProGlyGlyAspTyraAsnThrSerLeuPro----- 950
OY 5604 GTACGCTAACCCCTCAAAATGGTGCTGTAGGCAATGGGTCCGCAATATGCCCATCAGC 5663
Db 951 -----ArgProAlaLeuGly-----GlySerIleProThrLeuProLeuAr 964
OY 5664 CAGCGACGGTACTCTCTGGAATGCCAGGATTGATGGCGGACGAGACCGGAGGTATGCT 5723
Db 964 gSerAsnSerIleProGlyAlaArgProValLeu----- 975
OY 5724 AATGAATTCTTCGGAGAGCAACCAAGATCACAACAAATCCTGGG----- 5775
Db 976 -----GlnGlnGlnGlnGlnMetLeuGlnMetArgProGlyGluIlePr 990
OY 5776 -GCAAGCAATGCTATTAACTTCTTTCAGAAATTGCAATCAAAATGCTATTGTCAGCAGA 5834
Db 990 oMetGlyMetGlyAlaAsnProTyra-----GlyGlnAlaAlaAlaSerAsnGlnLe 1007
OY 5835 GGGTGATTAACCCCGGCATGACGATCAATGAAT----- 5868
Db 1007 uGlySerTrpPro-----AspGlyMetLeuSerMetGluGlnValSerHisGlyThrG 1025
OY 5869 -----ATTGGTCAACCATCTATGATATAAG 5891
Db 1025 nAsnArgProLeuLeuArgAsnSerLeuAspAspLeuValGlyProProSerAsnLeuG 1045
OY 5892 GGGCATG-----CGTCCACATGCCATGCGGCCAAATGTAAT 5927
Db 1045 uGlyGlnSerAspGluArgAlaLeuLeuAspGlnLeuHisThrLeuLeuSerAsnThrAs 1065
OY 5928 GGGTGCGCGATGCCACCCGTTAACAGGCAATT----- 5961
Db 1065 pAlaThrGlyLeuGlnGluIleAspArgAlaLeuGlyIleProGluLeuValAsnGlnG 1085
OY 5962 -----CAGTTTGACACAGTCATCGATGGATGTGACTGTGTCGGGA 6002
Db 1085 yGlnAlaLeuGluProLysGlnAspAlaPheGlnGlyGlnGluAlaAlaValMetMetAs 1105
OY 6003 TCCGTTCATCATTTTCACTAAGCTTCTGCAACAGCGCTGACACACATGTTGATC 6062
Db 1105 pGlnLysAlaGlyLeuTyrglyGlnThrTyrrProAlaGlnGlyProPrometGlnGlyG 1125
OY 6063 AGCACA---CAGGCCAATCAGCCTTAAGACACACACATAAAGAACATACCTAGTGAAT 6119
Db 1125 yPheHisLeuGlnGlnGlyGlnSerProSerPheAsnSerMetMetAsn----- 1140
OY 6120 GTGTCAAAACCAATCGGACTTGACGTGGCAACAAGG----- 6156
Db 1141 -GlnMetAsnGlnGlnGlyAsnPheProLeuGlnGlyMetHisProArgAlaAsnIleMe 1160
OY 6157 -----CAGATCCAACCTGATGGCAAGACATGCGCAGGG 6191
Db 1160 tArgProArgThrAsnThrProLysGlnLeuArgMetGlnLeuGlnGlnArgLeuGlnG 1180
OY 6192 TCAGTCTTAATTGAGCTACTATAATAATTAA-----ATGTCAACTGC 6236
Db 1180 yGlnGlnPheLeuAsnGlnSerArgGlnAlaLeuGluLeuLysMetGluAsnProThrAl 1200
OY 6237 CGGAAGTGTCAGTGCT-----ACTAACGGTGTCTCTGGCATCAA 6275
Db 1200 aGlyGlyAlaAlaValMetArgPrometMetGlnProGlnGlnGlyPheLeuAsnAlaG 1220
OY 6276 TTTGTAAGTCCCTCTTCTACGCACTG----- 6303
```

```
Db      1220 mMetValAlaGlnArgSerArgGluLeuLeuSerHisHisPheArgGlnGlnArgValAl 1240
Qy      6304 ----AAGTATGCCCAGCAATATCATAGTTTCACGACGACGTATATAGCTAACACACCAG 6359
         |||||
Db      1240 ametMetmetGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1260
Qy      6360 AAGTCAACAACAACAGCATATGCAACCAGCAGCACCGAACATGATAACAATGCCGCC 6419
         :|||||
Db      1260 nGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnThrglnAlaPheSerProProPr 1280
Qy      6420 GAATTATCACCCAATCCAACG 6441
         |||:::
Db      1280 cAsnValThrAlaSerProSer 1287
```

RESULT 14
US-09-949-016-9675

```

: Sequence 9675, Application US/09949016
: Patent No. 6812339
: GENERAL INFORMATION:
: APPLICANT: VENTER, J. Craig et al.
: TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
: TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
: FILE REFERENCE: CL001307
: CURRENT APPLICATION NUMBER: US/09/949,016
: CURRENT FILING DATE: 2000-04-14
: PRIOR APPLICATION NUMBER: 60/241,755
: PRIOR FILING DATE: 2000-10-20
: PRIOR APPLICATION NUMBER: 60/237,768
: PRIOR FILING DATE: 2000-10-03
: PRIOR APPLICATION NUMBER: 60/231,498
: PRIOR FILING DATE: 2000-09-08
: NUMBER OF SEQ ID NOS: 207012
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 9675
: LENGTH: 2451
: TYPE: PRT
: ORGANISM: Human
US-09-949-016-9675

```

Alignment Scores:

| | | | |
|------------------------|---------|---------------|------|
| Pred. No.: | 4.8e-11 | Length: | 2451 |
| Score: | 219.00 | Matches: | 297 |
| Percent Similarity: | 31.80% | Conservative: | 223 |
| Best Local Similarity: | 18.17% | Mismatches: | 557 |
| Query Match: | 1.82% | Indels: | 560 |
| DB: | 4 | Gaps: | 75 |

US-10-664-859-1 (1-6909) x US-09-949-016-9675 (1-2451)

| | | | |
|----|------|---|------|
| QY | 2249 | ATATATAGAAGTCGCGCTTCACACTCTTCTGGCGCGCTTCACCACTACGTGAGATCCGCC | 2308 |
| | | ::: ::: ::: | |
| Db | 441 | ValTyrIlysAspArg-----GlnPheMetAsnValTrpThrAspHis | 454 |
| QY | 2309 | CGCAGTGAATTATATAGA-----TGATTTACG | 2335 |
| | | :::~::~~::~~::~~::~~::~~:: | |
| Db | 455 | GluIysGluIlePheIysAspIysPheIleGlnHisProIysAsnPheGlyLeuIleAla | 474 |
| QY | 2336 | AGTTATTTTA-----ATTTTTATGCTGTATTTTAATAATA | 2371 |
| | | | |
| Db | 475 | SerTyrIleuGluArgIysSerValProAspCysValLeuTyr----- | 488 |
| QY | 2372 | TCTTATTATTATTCATTTTACATAGTTAAATTTGAAGAATTTCAACGACAGTACCACGGA | 2431 |
| | | ::: | |
| Db | 489 | ---TyrTyrIleuThrIlylSbIysAsnGluAsnTyrIlysalaleuValArgArgAsnTyrGly | 507 |
| QY | 2432 | AAAAAAGGATCGTCCTTGACAATGAATAATGACGAATGAGCATGGAAGGCTGCAATCA | 2491 |
| | | : | |
| Db | 508 | LysArgArg-----GlyArgAsnGlnGlnIleAlaArgProSer | 520 |
| QY | 2492 | GTTGAATCCCGATTTTATCAATGAATCTTTAAATAATCCTGCAATTTGAGCATATTAGT | 2551 |
| | | ::: | |
| Db | 521 | GlnGluGluIysValGluGlu-----LysGluGluAspIysAlaGluIysThrGlu | 537 |


```
Db      873  LAlaGlnGlnIleAsnAlaGlnArgProGluProGlnSerAspAsnAspSerSerAlaThr 893
QY      3428 GGGTGGCGGAGCGGATCTATTATTAATCCGACGATGCGACAACCTGCATATGCCAGGTAA 3487
Db      893  rCysSerAlaAspGluAspValAspGlyGluProGluArgGlnArgMetPheProMetAs 913
QY      3488 CGCCAATCGGAGCTCTTATCGGCGACAAGTTCAGACTTTCGGA----- 3533
Db      913  pSerLysProSerLeuLeuAsnProThrGlySerIleLeuValSerSerProLeuLysPr 933
QY      3534 -----GATGTAATGCATCCAGGGGATGTTATATCAGATATGGTGC 3574
Db      933  oAsnProLeuAspLeuProGlnLeuGlnHisArgAlaAlaValIleProPheMetValSe 953
QY      3575 CGTAATAGGATGTAATAATAATCAAAAAACAGTGCATATGTGATCTGAGTAGGTGT 3634
Db      953  rCysThrProCysAsn----- 959
QY      3635 TGTCACTGGAACAACCTGCAGCTGAGTA----- 3662
Db      959  eProIleGlyThrProValSerGlyTyrAlaLeuTyrGlnArgHisIleLysAlaMetHis 979
QY      3663 -----AATGCAATATGCATTGCTC 3682
Db      979  sGlnSerAlaLeuLeuGluGluGlnArgGlnArgGlnGlnIleAspLeuGluCysArg 999
QY      3683 AAGCTCCGGCGCCCGCAATGCGAATATGATGGGAAGCTCTACGGATATGCCTCGTT 3742
Db      999  gSerSerThrSerPro----- 1004
QY      3743 TGGCAACACAAAGCTGCAACGTCATCGGAACGCCCCAGATATGTCTAAGAAAGTTTAA 3802
Db      1005 -----CysGlyThrSerLysSerProAsnArg----- 1013
QY      3803 TCAAGATAGCCGAACCATTCATCATCAAGGGGAGTTGCTCAATGAGTGTGGAAGAT 3862
Db      1014 -----GluTrpGluValLe 1018
QY      3863 TCAA-----CATCAATTTTTCGAAGA-----CGCCTCAAGGGGG 3898
Db      1018  uGlnProAlaProHisGlnValIleThrAsnLeuProGluGlyValArgLeuProThrTh 1038
QY      3899 CAAGCCACAGACGCTCACTGGAAGTGTAGTACCAACAAGCAAAACC----- 3944
Db      1038  rArgProThrArgProProProProLeuIleProSerSerLysThrThrValAlaSerG1 1058
QY      3945 -----CCTTCTGATCTGTGGAAGTCTGTTAAACAACAGGTGCGACCCCTGCAAGTCC 4000
Db      1058  uLysProSerPheIleMetGlyGlySerIleSer-----GlnGlyTh 1072
QY      4001 ACCTCCTCTTACCACTCCATCCAGAGATGTGGTCAAGTACCAATAGCCACTCAATCGCC 4060
Db      1072  rProGlyThrTyrLeuThrSerHisAsnGlnAlaSer-----TyrThrGlnGluThrPr 1090
QY      4061 CAATCCCTCG-----AGTCCAAACAA 4081
Db      1090  oLysProSerValGlySerIleSerLeuGlyLeuProArgGlnGlnIleSerAlaLysSe 1110
QY      4082 TCTATCTCTCCG-----TCACGCGG----- 4103
Db      1110  rAlaThrLeuProTyrIleLysGlnGluLysPheSerProArgSerGlnAsnSerGlnPr 1130
QY      4104 -----ACAACCGCAGCAGT 4117
Db      1130  oGluGlyLeuLeuValArgAlaGlnHisGluGlyValValArgGlyThrAlaGlyAlaI1 1150
QY      4118 CATG-----GGATTGCCGCAACTCTCTAGCATGAGTGAACAGG 4159
Db      1150  eGlnGluGlySerIleThrArgGlyThrProThrSerLysIleSerValGluSerIlePr 1170
QY      4160 ATCATTATCTGATCTGTCCGCAAGCTAATACTTCAGCGGTTCAAGGAGGACAAACAAC 4219
```

```
Db      1170  oSerLeuArgGlySerIleThrGlnGlyThrProAlaLeuProGlnThrGlyIleProTh 1190
QY      4220 AGTGCTCTCAGCAACACAAAGACTGTTTTCAGGACAGACCCCATCGCCGTCAATCAAA 4279
Db      1190  r----- 1190
QY      4280 TCGTAGTAGAAATACCGGATCGTCAAGCGTCTTACCGCATTAATTAGCAGCAACCCAAG 4339
Db      1191  -GluAlaLeuValLysGlySerIleSerArgMet----- 1201
QY      4340 TACCCCTTATCTCATCTATCCCAAGAAAGAAATTTAGTCTTTCGGTCACTCTGCTGG 4399
Db      1202  -ProIleGluAspSerSerProGluLysGlyArgGluGluAlaAlaSerLysGlyHis 1220
QY      4400 TATGTTATATTGTTTAAATTTTAAAGCAAAATCAAAATGAATTCGTTAATGAATAA 4459
Db      1220  sValIleTyr-----GluGlyLysSer----- 1227
QY      4460 GTTATATATTACATACTCGAAATTTGATA-----GAAAAATCAGGATAGAAAATA 4513
Db      1228  -GlyHisIleLeuSerTyrArgAsnIleLysAsnAlaArgG1 1241
QY      4514 TAAATTATTTTC---CGAACCGCCCATCCATTCTTGAATCCAATTTCTGAGTGATTG 4569
Db      1241  uGlyThrArgSerProArgThrAlaHis----- 1250
QY      4570 TTAGAGATAATCTACTATTAAATTAACACGAAATTCATATCCGTTAATGAAATCA 4629
Db      1251  -GluIleSerLeuLysArgSerTyrGluSerValGluGly-- 1263
QY      4630 CTATTGTTTAATAAGAAATTAATAATATGTTTATTATTAATATTCTACAGTGATAACAT 4689
Db      1264  -AsnIleLysGlnGlyMetSerMe 1271
QY      4690 GAAAGTAGGCGCAGCAAGCCACAGGGTCAAGGTCACCAAGTAATAGTCTAATAGAGGC 4749
Db      1271  rArg-----GluSerProValSerAlaProLeuGluG1 1282
QY      4750 AAATTAAGATGTACGATTGTGTCGATCCAGTCTGTGTTTAAACCGCATCCACATATGCA 4809
Db      1282  y-----LeuIleCysArgAlaLeuProArgGlySerProHisSerAspLeuLys 1298
QY      4810 AAGCAAT-----TCAATTCAGCATTTAAACGCGCTATAAATGGGCTTACCACAA--- 4857
Db      1298  sGluArgThrValLeuSerGlySerIleMetGlnGlyThrProArgAlaThrThrGluSe 1318
QY      4858 -TATACAGATGAGGTAAATATTTAATATTATTATTAAACGTTTGTGTTAATTATCT 4916
Db      1318  rPheGluAspGlyLeuLysTyrProLysGlnIle----- 1329
QY      4917 TCTTTTCAAGCTCAAGCA-----TCAGCGCAAGGTGATCCGTAATTAAGTTCG 4967
Db      1330  -LysArgGluSerProProIleArgAlaPheGluGlyAlaIleThrLysGlyLys 1347
QY      4968 GCGTCCGATAATATTCCGCTAAATCCCAATAGTGAATCGG-----CCGCC 5015
Db      1347  sProTyrAspGlyIleThrThrIleLysGluMetGlyArgSerIleHisGluIleProArg 1367
QY      5016 ACCAAACAAGATGACCCCAAACTGCAATCCATCTCTTGGCACAATGTCCCAACA 5075
Db      1367  gGlnAspIleLeuThrGlnGluSerArgLysThrProGluValValGlnSerThrArgPr 1387
QY      5076 ACTAACAAGTTGCGGTGCCAGCATGGGTAGTCCA-----GCCGGA-- 5115
Db      1387  oIleIleGluGlySerIleSerGlnGlyThrProIleLysPheAspAsnAsnSerGlyG1 1407
QY      5116 -----ACTGCTGTATGACGATGATGGGGG 5141
Db      1407  nSerAlaIleLysHisAsnValLysSerLeuIleThrGlyProSerLysLeuSerArgG1 1427
QY      5142 TCCGGAGCGTCCGACATCAATATTTGACATGGAATATTTCCGGAATAAGATGATCAGG 5201
Db      1427  yMetProProLeuGluIle--ValProGluAsnIleLysValValGluArgGlyLysTy 1446
```

QY 5202 AATAGATACCATAAATCAATACTGTCAATTCATGAATGTCGTAATGAACATGCGG 5261
Db 1446 rgluAspVallysalaglyluThrValArgSerArgHisThrSerValValSerSergl 1466
QY 5262 TCCCCGAATGCTGAATCCTAAATGTGCGTAGCAGCGCGTCCAATGAGCCCGCTGCTT 5321
Db 1466 yProSerValLeuArgSerThrLeuHisgluAlaProlysalaglInLeuSerProglyl1 1486
QY 5322 TAATCCTAATTCGCCCAATGGTGGATTAGAGAGAAATTCATAGGGTCTGGCTGGCTC 5381
Db 1486 eTyraSpasPThr-----Se 1491
QY 5382 AGCAAACTCTTCA-----AACTTTCAGGGGGTGTTCACCTGTGCGCAGAAATGAGG 5435
Db 1491 rAlaArgArgThrProValSerTyrglnAsnThrMetSerArgglySerProMetMetAs 1511
QY 5436 TCGA--ATGCCAGTCAATTTGGTTGGAATTTCAATCCGAATATTCCAGTAAGCGAG 5492
Db 1511 nArgThrSerAspValThrIleSerSerAsnlySerThrAsnHisgluArglySerTh 1531
QY 5493 T---ACCCCAACACCATACATACATGCCAGTAGAGGACAGAACGCCAAC----- 5541
Db 1531 rLeuThrProThrGlnArggluSerIleProAlalySerProValProglyValAspPr 1551
QY 5542 -----AACAAATAAACAATGAGTGAATTAATGTG-- 5571
Db 1551 oValValSerHisSerProPheAspProHisArgglySerThrAlaglyluValTy 1571
QY 5572 -----CGAATGCCACTAGTCTGGAATTTTTCAGAGGTACGCTAA 5612
Db 1571 rArgSerHisLeuProThrHisLeuAspProAlaMetProPhe--HisArgAlaLeuAs 1590
QY 5613 CCCTCAATGGTGTGTAGGCAATGGTGCAGCAATATGCCCAATCAGCCAGCGACGCG 5672
Db 1590 pProAlaAlaAlaAlaAlaTyrluPheglInArgglInLeuSerPro----- 1604
QY 5673 TACTCCTGGAATGCCAGA-----TTGATGGCGGAGACCAGAGCCGAGTATGCT 5723
Db 1605 -ThrProglyTyrlProSerGlnTyrglnLeuTyrlaMetGluAsnThrArgglInThr1 1624
QY 5724 AATGAATTTCTCCGAGAGCAACACAGAACAGATCAACAACAATCCTGGGGCAAGCAA 5783
Db 1624 eleuAsnAspTyrlIleThrSerGlnGlnMetGlnValAsnLeuArgProAspValAlaAr 1644
QY 5784 TGGTATTAACTTTCTGAAATTGCAATCAATGTCTATTGTCAGAGAGAGGTGAT 5843
Db 1644 gGly-----leuSerProArggluGlnProleuGlyle 1655
QY 5844 ACCCGGCATGACGGATCAATGATATTGTCACCATCTATG-----AT 5888
Db 1655 uPro-----TyrProAlaThrArgglyIleleAspLe 1666
QY 5889 AAGGGCATGCGTCCACATGCCATGCGGCGCAATGTAATGGGTGCGCGATGCCACCGT 5948
Db 1666 uThrAsnMetProProThrIleleuValProHisProglyTyrlThrSerThrProProme 1686
QY 5949 TAACAGG-----CAATTCACTTTGCACAGTCATCGAGTGTAT 5987
Db 1686 lAspArgIleThrTyrlleProglyTyrlGlnIleThrPhe----- 1699
QY 5988 TGACTGTGTGGGATCGGTCAATCTTTTCACTTAACGCTTCTGCAACAGCGGTGACC 6047
Db 1700 -----ProProArgProTyrlAsnSerAlaSerMetSerProglyHisPr 1714
QY 6048 A---CACATGTTTGATGACGACCAACAGCGCAATCAGCCTTAAGACACAAACATAGA 6104
Db 1714 cThrHisLeuAlaAlaAlaAlaSerAlagluArggluArggluArggluArggluysgl 1734
QY 6105 CATACCTAGTGAATGTCAAAAACAATCGGAGCTTGACGTGACCAAGGCGAGATCCA 6164
Db 1734 u-----ArggluArggluArggluAlaAlaAlaAlaSerSerAspLeuTy 1748

QY 6165 ACTGCAT-----GGGCAAGACATGCGCAGGCTCA 6194
Db 1748 rleuArgProglySerGlnGlnProglyArgProglySerHisglTyrlValArgSerPr 1768
QY 6195 GTCTTTAATGGACTACTAATAATTAATTAACTGCAACTGCCGAGAGTGC--AGTGC 6251
Db 1768 oSerProSerValArgThrGlnGlnIleuThrMetleuGlnGlnArgProSerValPheglIngl 1788
QY 6252 TACTAAGGTGTCTGTGCATCAATTTGTAAGTCCCTCTTCT 6294
Db 1788 yThrAsnGlyThrSerValIleThrProleuAspProThrAla 1802

RESULT 15

US-08-701-154A-5
; Sequence 5, Application US/08701154A
; Patent No. 6380373
; GENERAL INFORMATION:
; APPLICANT: O'Malley, Bert W.
; APPLICANT: Tsai, Ming-Jer
; APPLICANT: Tsai, Sophia Y.
; APPLICANT: Onate, Sergio A.
; TITLE OF INVENTION: STEROID RECEPTOR COACTIVATOR
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: FastSeq for Windows 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/701,154A
; FILING DATE: August 21, 1996
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/003,784
; FILING DATE: September 15, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 220/243
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1061 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-701-154A-5

Alignment Scores:

Pred. No.: 5.05e-11 Length: 1061
Score: 216.50 Matches: 239
Percent Similarity: 31.28% Conservative: 152
Best Local Similarity: 19.12% Mismatches: 410
Query Match: 1.80% Indels: 449
DB: 3 Gaps: 58

US-10-664-859-1 (1-6909) x US-08-701-154A-5 (1-1061)

QY 2952 TTGAACCCCAATTCATCAATCAACATGAAGAAACAGCAAACTAGTACTGTAAAGCGCGCTAGC 3011
::: ||| :::: |||||:::

Db 10 ValAsnProSerIleSerProAlaHisGlyValAlaArgSerSerThrLeuProProSer 29
QY 3012 AACTCTTTTGCAGACCAGTGTGATCCTATGGCAACGAACCTGAATGATGTGTGGAA 3071
Db 30 AsnSerAsnMet----- 33
QY 3072 GCGCGATCCTCAAAACACCAGTAGCTGTGACAAAACCTCAGCAATCATGTAGACAGTATC 3131
Db 34 --ValSerThrArgIleAsnArgGln---GlnSerSerAspLeuHisSerSerHis 51
QY 3132 AGTACATCCAGCGAGTCAACAG-----GCAATAAAGATACTGGAA 3170
Db 52 SerAsnSerSerAsnSerGlnGlySerPheGlyCysSerProGlySerGlnIleValAla 71
QY 3171 GCAGCTGCGTGTGATTTGGACAGGTC---ACAAAGAGAGCGATCCTGGCCTGACAACT 3227
Db 72 AsnValAlaLeuAsnLysGlyGlnAlaSerSerGlnSerSerLysProSerLeuAsnLeu 91
QY 3228 GAAACAACACATTGTATCACTGCAGAGAGTTAAGTTCAGACGAACAACTTACACCACAA 3287
Db 92 AsnAsnProPrometGluGlyThrGlyIleSerLeuAla---GlnPheMetSerPro--- 109
QY 3288 CAGCGGCACATCGGGAAGAACAGTTCGCAAAAAATAAAAAATGAATCAATTTCTTTT 3347
Db 110 --ArgArgGlnValThrSerGlyLeuAlaThrArgProArgMetProAsnAsnSerPhe 128
QY 3348 CCTGAAATGAGAAATTCAGTAGAGCT-----AAT 3377
Db 129 ProProAsnIleSerThrLeuSerSerProValGlyMetThrSerSerAlaCysAsnAsn 148
QY 3378 GTAAGCTCAGATTAACAAAAATTCAGAGAGATTATGATGGGATGTCGGGTGCGGA 3437
Db 149 AsnAsnArgSerTyrrSerAsnIleProValThrSerLeuGlnGlyMetAsnGluGlyPro 168
QY 3438 GCGCGATCTAT-----ATAATCCGACGATCGACAACTGCAT----- 3476
Db 169 AsnAsnSerValGlyPheSerAlaSerSerProValLeuArgGlnMetSerSerGlnAsn 188
QY 3477 -----ATGCCAGTAAACGCCCAATCG----- 3497
Db 189 SerProSerArgLeuAsnIleGlnProAlaLysAlaGluSerLysAspAsnLysGluIle 208
QY 3498 -----GAGCTCTTATCGGCCACAGTTCAGGA-----CTT 3527
Db 209 AlaSerThrLeuAsnGlnMetIleGlnSerAspAsnSerSerSerAspGlyLysProLeu 228
QY 3528 TCGGAAGATGTAATGCATCCAGGGAGTGTATATAGATATGGTCCGTAAATAGATGT 3587
Db 229 AspSerGlyLeuLeuHisAsnAsnAspArgLeuSerAsp-----GlyAsp 243
QY 3588 AATAATATCAAAAAACCAAGTGTGCAATGTGCATCTGAGTAGAGTGTGTCACTGGAACA 3647
Db 244 SerLysTyrrSerGlnThrSerHisLysLeu-----ValGlnLeuLeuThr---Thr 259
QY 3648 ACTGCA-----GCTGAGTAATGTCAATATGCAATTGC-----TCA 3683
Db 260 ThrAlaGluGlnGlnLeuArgHisAlaAspIleAspThrSerCysLysAspValLeuSer 279
QY 3684 AGCTCCGGCGCCCGAATGGCAATATGATGGAGAACTCTACGGATATGTAAGCTCGTTT 3743
Db 280 CysThrGlyThrSerAsnSerAlaSerAlaAsnSerSerGly----- 293
QY 3744 GGCACACACAGCTGCAACGTCACTCGAACGGCCCCAGATATGTCTAAGGAAGTTTAAAT 3803
Db 294 -----GlySerCysProSer-----Ser 299
QY 3804 CAAGATAGCCGAACCCATTACATCAAGGGGAGTTGCTCAATGAGTGTGAGAGATT 3863
Db 300 HisSerSerLeuThrAlaArgHis-----LysIle 309
QY 3864 CAACATCAATTTTGAAGAAGCGCTCAAGGGGGCAAGCCGACAGACAGTCACTGGAAGT 3923
Db 310 LeuHisArgLeuLeuGlnGlu-----GlySerProSerAspIleThrThrLeu 325

QY 3924 GTAATACCACACAGCAAAACCCCTTCTGATCTGTGGAACCTGTTAAACAACCAAGTG 3983
Db 326 SerValGluProAspLysAspSerAlaSerThrSerValSerValThrGlyGlnVal 345
QY 3983 ----- 3983
Db 346 GlnGlyAsnSerSerIleLysLeuGluLeuAspAlaSerLysLysGluSerLysAsp 365
QY 3984 -----CGACCCCTGCAAGTCCACCT--- 4004
Db 366 HisGlnLeuLeuArgTyrrLeuLeuAspLysAspGluLysAspLeuArgSerThrProAsn 385
QY 4005 -----CCTCCTTACCACCTCCATCCAGAGATCTGCGTCACTACCAATAGCCACT 4052
Db 386 LeuSerLeuAspAspValLysValLysValGluLysLysGlnMetAspProCysAsn 405
QY 4053 CAATCGCCCAATCCC-----TCGAGTCCAAACAATCTATCTCTCCCGTCACCGCGG 4103
Db 406 ThrAsnProThrPrometThrLysAlaThrProGluGluIleLysLeuGluAlaGlnSer 425
QY 4104 ACAACCGCAGCAGTCATGGA-----TTGCGGACC----- 4133
Db 426 GlnPheThrAlaAspLeuAspGlnPheAspGlnLeuLeuProThrLeuGluLysAlaAla 445
QY 4134 AACTCTCCTAGCATGATGAGACAGATCATATATCTGATCTGTTCGCAAGCTAAT--- 4190
Db 446 GlnLeuProGlyLeuCysGluThrAspArgMetAspGlyAlaValThrSerValThrIle 465
QY 4191 -----ACTTCGACGGTTCAGGACGACGACACACACAGTGTCTCAGCAAC 4235
Db 466 LysSerGluIleLeuProAlaSerLeuGlnSerAlaThrAlaArgProThrSerArgLeu 485
QY 4236 AAGAACTGTTTTCAGGACAGACACCCCATCGCGTCAAAATCAAAATCGTAGTAAGAAATACC 4295
Db 486 AsnArgLeuProGluLeuGluLeuGluAlaIleAspAsnGlnPheGlyGlnProGlyThr 505
QY 4296 GGA-----TCGTCAAGCGTCTTACGCATACTTAAGCAGACCAACCA 4337
Db 506 GlyAspGlnIleProTrpThrAsnAsnThrValThrAlaIleAsnGlnSerLysSerGlu 525
QY 4338 -----AGTACCCCTTATCTCATCTA-----TCCCAAAAGAAATTGAGTCT 4379
Db 526 AspGlnCysIleSerSerGlnLeuAspGluLeuLeuCysProProThrThrValGluGly 545
QY 4380 TTCGGTCAGTCTCTGCTGTATGTATATTTGTTAATTTTTT-----AAAGACAAA 4433
Db 546 ArgAsnAspGluLysAlaLeuLeuGlnLeuValSerPheLeuSerGlyLysAspGlu 565
QY 4434 TCAATATGAATTGCGTTAATATTAAGTTATATATTTACATACTC-----GGAATTTG 4487
Db 566 ThrGluLeuAlaGluLeuAspArgAlaLeuGlyIleAspLysLeuValGlnGlyGly 585
QY 4488 ATGAAAAAAATCAGGAATAGAAAAATTAATTTTCCGACCGCCCATCATTTCTTG 4547
Db 586 LeuAspValLeuSerGluArgPheProProGlnGlnAlaThrProProLeu--IleMetG 605
QY 4548 AATCC-----AATTTCGAGTGATGTGTAGAGATAATCTAATTAAATTAACACAG 4601
Db 605 IuGluArgProAsnLeuTyrrSerGln-----ProTyrS 616
QY 4602 AAAATTCATATCCGTTAATTGAAATCACTATTGTTAATAGAAATTAAATATATGTTT 4661
Db 616 erSerProPheProThrAla----- 622
QY 4662 ATTATATATT---TCTACAGGTGATAACATGAAGAAAGTACGACCAAGC----- 4708
Db 623 -----AsnLeuProSerProPheGlnGlyMetValArgGlnLysProSerLeuGlyThrm 641
QY 4709 -----CCACAGGGTCAGCGTCAACCA----- 4729
Db 641 etProValGlnValThrProProArgGlyAlaPheSerProGlyMetGlyMetGlnProA 661

QY 4729 ----- 4729
 Db 661 rGlnThrLeuAsnArgProProAlaAlaProAsnGlnLeuArgLeuGlnGlnGlnA 681
 QY 4730 -----GTAATAAGTCTAATAGAGGCAATAAAGATGTA-----CGATTTG 4769
 Db 681 rGleuGlnGlnGlnGlnLeuLeuHISGlnAsnArgGlnAlaLeuAsnGlnPheA 701
 QY 4770 CTGCATCCAGTCCCT---GGTTTAAACCCGCATCCACATATGCAAGC-----AATTCAA 4820
 Db 701 laAlaThrAlaProValGlyIleAsnMetArgSerGlyMetGlnGlnGlnIleThrProG 721
 QY 4821 ATTCAGCATTAACCGCTATAAATAGGGCTCTACCAATATACAGATGGAGTAAATATT 4880
 Db 721 InProProLeuAsnAlaGlnMetLeuAlaGlnArgGlnArgGluLeu----- 736
 QY 4881 AAATATTATTATTAAAGTTTGTGTGTTAATTATCTTTTTCAGCGCTCAGCATCAGC 4940
 Db 737 -----TyrSerGlnGlnHisArgG 743
 QY 4941 GCAAGGTGATCCGTACAATTAGTCGGCGCTCGATAATATTCCGTTAAATCCCAATAG 5000
 Db 743 In-ArgGlnLeu-IleGlnGlnGlnArgAlaMet-----LeuMetArgGlnGlnSerPh 760
 QY 5001 TGGCAATCGCCGCCCAACAAGATGACCCAAACTTCGATCCAATCTCTTTTGGC 5060
 Db 760 eGlyAsnAsnLeuProPro----- 766
 QY 5061 ACAATGTCCCAACAATAACAAGTTCGCTGTCCAGCATGGGTAGTCCAGCCGAACTGG 5120
 Db 767 -----SerSerGlyLeuProValGlnThrGln 775
 QY 5121 TGGTATGACGATGATGGGGGCTCCGGGACCGTCCGACATCAATATTGAGCATGAATAT 5180
 Db 775 Y----- 775
 QY 5181 TTCGGAGTAGATGATCAGAATAGATACATAAATCAAAATACTGTCAATCAATGAA 5240
 Db 775 ----- 775
 QY 5241 TGTGTAATGAACCTCAATGGGTCCCGGAATGCTGAATCCTAAATGTGCGTAGCAGCGG 5300
 Db 776 -----AsnProArgLeu----- 779
 QY 5301 TCCAATGACCCGCT-----GGCTTAACTCTAATTC 5333
 Db 780 -ProGlnGlnAlaProGlnGlnPheProTyrProProAsnTyrGlyThrAsnProGlyTh 799
 QY 5334 CCCCATGTGTGATTAAGAGAATTCATAGGGTCTGGCTGTGCTCAGCAAACTCTTC 5393
 Db 799 rProProAlaSerThrSerProPheSerGlnLeuAlaAlaAsnProGlnAlaSerLeuAl 819
 QY 5394 AAACCTTCAAGGGGTGTTCACCTGTGTCAGAAATGATGGTCAATGCCAGTCAATTT 5453
 Db 819 aAsnArgAsnSerMetValSerArgGly-----MetThrGlyAsnIleGlyGlnPh 837
 QY 5454 TGGTTCGAATTTCAATCCGAATATTCAAGTAAGCGAGTACCCCAACACACATACAATA 5513
 Db 837 eGlyThrGlyIleAsnProGlnMetGln-----GlnAsnValPheGlnTy 852
 QY 5514 C-----ATGCCAGTAAGGGCACAGAACGCCCAACAACAATAACAATGG 5558
 Db 852 rProGlyAlaGlyMetValProGlnGlnGlyAlaAlaAsnPheAlaProSerLeuSerProGln 872
 QY 5559 AGCTAATATATGTCGAATG-----CCACCTAGTCTGGAATTTTTCAGAGG----- 5604
 Db 872 ySerSerMetValProMetProIleProProGlnSerSerLeuLeuGlnGlnThrPr 892
 QY 5605 -----TACGCTAACCT-----CAAATGGGTGCTGTAGGCAA 5636
 Db 892 oProAlaSerGlyTyrGlnSerProAspMetLeuAlaTrpGlnGlnGlyAlaIleGlyAs 912
 QY 5637 TGGGTGCGCAATATGCGCACCATCAGCCAGCGAGCGTACTCCTGAATGCCAGGATTG-- 5694

Db 912 nAsnAsnValPheSerGlnAlaValGlnAsnGlnProThrProAlaGlnProGlyValTy 932
 QY 5695 -----ATGCCGGGACACAGAGCCGAGGTATGCTAATGAA 5729
 Db 932 rAsnAsnMetSerIleThrValSerMetAlaGlyGlyAsnThrAsnValGlnAsnMetAs 952
 QY 5730 TTCTTCGGGAGGACAACACAGAGATCAACAATCCTGGGGCAAGC----- 5781
 Db 952 nProMetMetAlaGlnMetGlnMetSerSerLeuGlnMetProGlyMetAsnThrValCy 972
 QY 5782 -----AATGTATTAACTTCTTTCAGAAATG 5807
 Db 972 sProGlnGlnIleAsnAspProAlaLeuArgHisThrGlyLeu-----TyrCy 988
 QY 5808 CAATCAAAATGTCTATTGTTGAC-----GAAGAGGTGATTAACCGGCATGACGG 5858
 Db 988 sAsnGlnLeuSerSerThrAspLeuLeuYsThrGluAla-----AspGln 1003
 QY 5859 ATCAATGAATATTGTCACCATCTATG-----ATAAG 5891
 Db 1003 YThrGlnGlnValGlnGlnValGlnValPheAlaAspValGlnCysThrValAsnLeuVa 1023
 QY 5892 GGGCATCGGTCCACATGCCATGCGGCCCAATGTAAATGGGTGCGCGGATGCCACCCGTTAA 5951
 Db 1023 lGlyGlyAspProTyrLeuAsnGlnProGlyProLeuGlyThrGlnYsProThrSerGln 1043
 QY 5952 CAGGCAATTCAGTTTGACACAG 5973
 Db 1043 yProGlnThrProGlnAlaGln 1050

Search completed: November 2, 2005, 22:23:59
 Job time : 275 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: November 2, 2005, 22:16:48 ; Search time 821.5 Seconds
(without alignments)
6505.484 Million cell updates/sec

Title: US-10-664-859-1

Perfect score: 12037

Sequence: 1 acgagtgcttcctcttattat.....aaaaaaaaaaaaaaaa 6909

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2105692 segs, 386760381 residues

Total number of hits satisfying chosen parameters: 4211384

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool.p/US10664859/runat_02112005_171806_3593/app_query.fasta_1.7047
-DB=A_Geneseq -QFMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10664859_@CCG_1_1_995_@runat_02112005_171806_3593 -NCPUT=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_16Dec04:*

1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|------------|--------------------|
| 1 | 7257 | 60.3 | 1464 | 5 AAB71228 | Aab71228 D. melano |
| 2 | 7257 | 60.3 | 1464 | 7 ABW01527 | Abw01527 Drosophil |
| 3 | 7257 | 60.3 | 1464 | 8 ADJ71911 | Adj71911 Fruit fly |
| 4 | 7040 | 58.5 | 1429 | 4 ABB58779 | Abb58779 Drosophil |
| 5 | 570 | 4.7 | 112 | 7 ABW01532 | Abw01532 Drosophil |
| 6 | 570 | 4.7 | 112 | 8 ADJ71898 | Adj71898 Fruit fly |
| 7 | 321.5 | 2.7 | 1435 | 4 ABB11808 | Abb11808 Human BCL |
| 8 | 314 | 2.6 | 1394 | 8 ADQ18945 | Adq18945 Human sof |
| 9 | 312 | 2.6 | 1426 | 5 AAB71229 | Aab71229 Human leg |
| 10 | 312 | 2.6 | 1426 | 7 ABW01534 | Abw01534 Human lgs |

| | | | | | |
|----|-------|-----|------|-------------|---------------------|
| 11 | 312 | 2.6 | 1426 | 7 ADJ70152 | Adj70152 Human hea |
| 12 | 312 | 2.6 | 1426 | 8 ADJ71903 | Adj71903 Human lgs |
| 13 | 296.5 | 2.5 | 1494 | 5 AAU78460 | Aau78460 Mouse bet |
| 14 | 270.5 | 2.2 | 1596 | 6 AAE36095 | Aae36095 Drosophil |
| 15 | 269.5 | 2.2 | 1594 | 4 ABB63688 | Abb63688 Drosophil |
| 16 | 258.5 | 2.1 | 1366 | 4 ABB66921 | Abb66921 Drosophil |
| 17 | 258.5 | 2.1 | 1251 | 4 ABB60086 | Abb60086 Drosophil |
| 18 | 247.5 | 2.1 | 1902 | 6 ABBJ38695 | Abbj38695 Human nuc |
| 19 | 242.5 | 2.0 | 1161 | 4 ABBJ70667 | Abbj70667 Drosophil |
| 20 | 241.5 | 2.0 | 1778 | 4 ABB60081 | Abb60081 Drosophil |
| 21 | 241 | 2.0 | 2280 | 4 ABB61650 | Abb61650 Drosophil |
| 22 | 240 | 2.0 | 2112 | 4 ABB60403 | Abb60403 Drosophil |
| 23 | 233.5 | 1.9 | 1243 | 7 ADJ70158 | Adj70158 Human hea |
| 24 | 233.5 | 1.9 | 1441 | 7 ADC23487 | Adc23487 Receptor |
| 25 | 233 | 1.9 | 2703 | 4 ABB63299 | Abb63299 Drosophil |
| 26 | 233 | 1.9 | 2703 | 5 ABG70019 | Abg70019 Larval vi |
| 27 | 233 | 1.9 | 3190 | 4 AAB84634 | Aab84634 Amino aci |
| 28 | 231.5 | 1.9 | 1844 | 3 AAB18250 | Aab18250 Plasmodiu |
| 29 | 230.5 | 1.9 | 3275 | 4 ABB70437 | Abb70437 Drosophil |
| 30 | 230.5 | 1.9 | 3275 | 7 ADK11301 | Adk11301 Drosophil |
| 31 | 229.5 | 1.9 | 2005 | 6 ABG74435 | Abg74435 Human nuc |
| 32 | 228.5 | 1.9 | 2063 | 4 ABG17147 | Abg17147 Novel hum |
| 33 | 228.5 | 1.9 | 2063 | 6 ABG74436 | Abg74436 Human nuc |
| 34 | 228.5 | 1.9 | 2063 | 8 ADQ91924 | Adq91924 Human act |
| 35 | 228 | 1.9 | 1424 | 6 ABU03694 | Abu03694 Human exp |
| 36 | 228 | 1.9 | 1424 | 6 ABU03681 | Abu03681 Human exp |
| 37 | 228 | 1.9 | 1424 | 6 ABU03684 | Abu03684 Human exp |
| 38 | 228 | 1.9 | 1424 | 7 ADE76188 | Ade76188 Human AIB |
| 39 | 225.5 | 1.9 | 2063 | 4 AAM78710 | Aam78710 Human pro |
| 40 | 225.5 | 1.9 | 2063 | 4 AAM40064 | Aam40064 Human pol |
| 41 | 223.5 | 1.9 | 1522 | 2 AAY21975 | Aay21975 Human ste |
| 42 | 223.5 | 1.9 | 1522 | 6 ABU03689 | Abu03689 Human exp |
| 43 | 222.5 | 1.8 | 1063 | 5 ABB54168 | Abb54168 Lactococc |
| 44 | 222.5 | 1.8 | 1441 | 4 AAE12570 | Aae12570 Human ste |
| 45 | 222.5 | 1.8 | 1441 | 5 AAE25569 | Aae25569 Human ste |

ALIGNMENTS

| | |
|-------------|---|
| RESULT 1 | |
| AAB71228 | |
| ID AAB71228 | standard; protein; 1464 AA. |
| XX AC | AAB71228; |
| XX DT | 18-NOV-2002 (first entry) |
| XX DE | D. melanogaster lgs protein. |
| XX KW | Legless; fruitfly; lgs; Wnt/Wingless signaling pathway; Wnt; Wg; |
| XX KW | tissue proliferation; tumour; cytoslatic; cellular disorder; colon; |
| XX KW | blood disorder; cancer; breast; head and neck cancer; brain; thyroid; |
| XX KW | medulloblastoma; skin cancer; tissue regeneration; tissue repair. |
| XX OS | Drosophila melanogaster. |
| XX PN | US2002086986-A1. |
| XX PD | 04-JUL-2002. |
| XX PF | 27-JUL-2001; 2001US-00915543. |
| XX PR | 28-JUL-2000; 2000US-0221502P. |
| XX PA | (BASL/) BASLER K. |
| XX PA | (BRUN/) BRUNNER E. |
| XX PA | (FROE/) FROESCH B. |
| XX PA | (KRAM/) KRAMPS T. |
| XX PA | (PETE/) PETER O. |
| XX PI | Basler K, Brunner E, Froesch B, Kramps T, Peter O; |
| XX DR | WPI; 2002-635689/68. |

DR N-PSDB; AAF88466.
XX
PT Novel polypeptide useful in therapeutic method for treating disorders of
PT cell fate such as cell differentiation or cell proliferation.
XX
PS Example II; Fig 2; 41pp; English.
XX
CC This invention describes a novel polypeptide sharing one or more
CC homologous amino acid domains with the legless (lgs) protein, a
CC downstream component of the Wnt/Wingless (Wnt/Wg) signaling pathway
CC involved in the formation and maintenance of spatial arrangements and
CC proliferation of tissues during development, and in the formation and
CC growth of many human tumours. The products of the invention have
CC cytosolic activity and can be used to treat cellular disorders, blood
CC disorders and cancers caused by over-stimulation of the Wnt pathway,
CC where the cancerous condition is colon, breast, head and neck, brain,
CC thyroid, medulloblastoma or skin cancer. The product could also be used
CC to promote tissue regeneration and repair. This sequence represents the
CC Drosophila melanogaster (fruitfly) legless (lgs) protein described in the
CC disclosure of the invention
XX
SQ Sequence 1464 AA;

Alignment Scores:
Pred. No.: 0 Length: 1464
Score: 7257.00 Matches: 1460
Percent Similarity: 73.01% Conservative: 1
Best Local Similarity: 72.96% Mismatches: 3
Query Match: 60.29% Indels: 537
DB: 5 Gaps: 5

US-10-664-859-1 (1-6909) x AAB71228 (1-1464)

QY 468 ATGCCCGCAGTCCAAACCAACAGCGCAACCAACTCCGATGCTCTCAACAAGT 527
DB 1 MetProArgSerProThrGlnGlnGlnProAsnSerAspAlaSerSerThrSer 20
QY 528 GCATCTGATCAAAATCCTGGAGCAGCGATCGGAAATGGGACTCGCGCGAGCAGAAGT 587
DB 21 AlaSerGlySerAsnProGlyAlaAlaIleGlyAsnGlyAspSerAlaAlaSerArgSer 40
QY 588 TCTCCGAAGACCTTATATAGCGAACCTTTTCTACTTTGTCCGCCGGTAAGACTTGATTT 647
DB 41 SerProIySerThrLeuAsnSerGluProPheSerThrLeu-Ser----- 54
QY 648 GATTTCTCTTTGTCGGATTTATACACTTTCTGTGTTCCAGATCAAAATTAATTTGAC 707
DB 55 -----ProaspGlnIleLysLeuth 61
QY 708 GCCAGAGAAGGCACTGAGAAAAAGCGGACTATCACTAGTATAAAGCTGCCACTGGAGG 767
DB 61 rProGlnGlnGlyThrGlnLysSerGlyLeuSerThrSerAspLysAlaIleThrGlyG1 81
QY 768 AGCCCCAGGAGTGGAATATATCTGCCGAGGAGCAAACTATGCTAAGGAGCAACTCTAC 827
DB 81 yAlaProGlySerGlyAsnAsnLeuProGlnGlyGlnThrMetLeuArgGlnAsnSerTh 101
QY 828 GAGCAACAATCACTCGTGCCTAGTGCCTTCTCCAAAACTCCAGTGAACACTCGAATAG 887
DB 101 rSerThrIleAsnSerCysLeuValAlaSerProGlnAsnSerSerGluHisSerAsnSe 121
QY 888 CAGCAATGTGTCTCTACAGTGGGCTTACTCAGATGTAGATTGTGACGAGCAATCGAA 947
DB 121 rSerAsnValSerAlaThrValGlyLeuThrGlnMetValAspCysAspGlnGlnSerLy 141
QY 948 GAAAAACAATGTAGTGTGAAGGAGGAGAACTGTGTAGACTGCCCTACAATGTTTA 1007
DB 141 sLysAsnLysCysSerValLysAspGlnGlnAla----- 152
QY 1008 AAATTTTAAATGTATTGGCGTTACCTTTGTTAATCATTTAATGTTTTTTTTTGCTA 1067
DB 152 ----- 152

QY 1068 TACTTACAATTTAGTTTAAACTGTAAACTTGACTTAAAACTCGCGAAGCTCGATCAA 1127
DB 152 ----- 152
QY 1128 AACAGACATTTTCTTGAACCGTAATTAAGCTCATATAAAATATTATTCATCTTGATGA 1187
DB 152 ----- 152
QY 1188 ATGCATATCATAGATGTACTCAACATCTCAAGAAAGACCTCAAAATTGGATCAACTAATT 1247
DB 152 ----- 152
QY 1248 AGTTTGAGAAAAAATTGCTGTACTTTTAAGATATATTAAATTTAAAAATTGCTGAGTGA 1307
DB 152 ----- 152
QY 1308 AATGATATAATAGTCACAATAATTTTAGTTAACTGCTAAAGCATTTTGATAGCCGTG 1367
DB 152 ----- 152
QY 1368 CTACGAGATGCTACTAGACGGGTGTAAGCTAATTTTATTAAAGCTGTCTTAAT 1427
DB 152 ----- 152
QY 1428 ATTCCATAACCATTAATGTCCATTTCAGAAATTAAGTTCTAATTAAGCAAAAGGTCAAGC 1487
DB 153 -----GluIleSerSerAsnLysAlaLysGlyGlnAl 163
QY 1488 AGCTGTGCGCGCTGCGAAACAGTTCTTACATCCAGTTTGACTGTCAAGAGAAACCAC 1547
DB 163 aAlaGlyGlyGlyCysGluThrGlySerThrSerSerLeuThrValLysGluGluProTh 183
QY 1548 CGATGCTTAGGCACTTTAGTAAATATGAAGAAAAAGAGAAAGAAATCATTCGCCAAC 1607
DB 183 rAspValLeuGlySerLeuValAsnMetLysLysGluIleArgGluAsnHisSerProTh 203
QY 1608 GATGTCCTCTGTTGTTTGGTTCAATTGGTAATGCACAGACAACTCCGCTACACCCGG 1667
DB 203 rMetSerProValGlyPheGlySerIleGlyAsnAlaGlnAspAsnSerAlaThr----- 221
QY 1668 TAAGTTTAAAGAGATCCATATATAAGCAAAATTAACAAGAAATTAATGTCAGTTACCAATTTTA 1727
DB 221 ----- 221
QY 1728 TTTGATAGTCAAGAAGACTACTATAGCGATATCTCGCTTTTAAATTTTATTTAATTAG 1787
DB 221 ----- 221
QY 1788 GAAATACGAATATTTCTAATTTGTAAATAAATAATTGATTAATTAAGATTTAAAAAC 1847
DB 221 ----- 221
QY 1848 CTTTGAATTAAGACATACCCTTCCAAAAATCAGTAATCATTTGGGAACGAGAGTGTGTC 1907
DB 221 ----- 221
QY 1908 CCGAAGAGACTACTATATAAACCTTTTGAAGCTATCTGATACTGACGCTACTAAAAATGA 1967
DB 221 ----- 221
QY 1968 TTAGTTTAGAAAAATGGGTGTAATTTTGTAGGAAGTTTCAATTTTAGAAGAAATGTGATT 2027
DB 221 ----- 221
QY 2028 ATTTTATTAACCCCTTCAAGCGAACTACATTGTTCTACGATATTTTGAAAAACAACA 2087
DB 221 ----- 221
QY 2088 TGGTTAAGTTGGAAGTGCCCTATTAACAAGAAATTCACGGTTTCAAAATACTAACACAGTT 2147
DB 221 ----- 221
QY 2148 TTTGATTTAATTTTGATTAATGAGAAATTTATCACACTTCAGTTAAATGTTAATTGCA 2207

Db 221 ----- 221
QY 2208 TTAAGTCGACAATCA CAGCAGATTTCATTTTGGCGTATATATAGAAGTCGCTTC 2267
Db 221 ----- 221
QY 2268 ACACTCTTGGCGGCTTCACCACCTAGCTGGAGTCCGCCCGCAGTGATTATATAGAT 2327
Db 221 ----- 221
QY 2328 GATTTCAGGTTATTATTTTATGTTGTTATTTAATAATATCTTATTATTTCATTT 2387
Db 221 ----- 221
QY 2388 TACATAGTTAAATTTGAAGAATTTCAAACGACAGTACCA CGGAAAAAAGATCGTCC 2447
Db 222 -Pro--ValLysIleGluArgIleSerAsnAspSerThrThrGluLysLysGlySerSer 240
QY 2448 TTGACAATGAATAATGACGAATGAGCATGGAAGGCTGCAATCAGTTGAATCCGATTTT 2507
Db 241 LeuThrMetAsnAsnAspGluMetSerMetGluGlyCysAsnGlnLeuAsnProAspPhe 260
QY 2508 ATCAATGAATCTTTAATAATATCCTGCAATTCGAGCATATTAGTAAGCGGAGTAGAACCA 2567
Db 261 IleAsnGluSerLeuAsnAsnProAlaIleSerSerIleLeuValSerGlyValGlyPro 280
QY 2568 ATACCCGGAATCGAGTTGGAGCGGGGACGGGAAATTTATGACTGCCAACGCCAATGGA 2627
Db 281 IleProGlyIleGlyValGlyAlaGlyAlaGlyThrGlyAsnLeuLeuThraIaAsnIaAsnGly 300
QY 2628 ATCTCCTCGGAGTAGCAGTAATTGTTGGATTACATGCAACAGCAAAATCACATATTCGTG 2687
Db 301 IleSerSerGlySerSerAsnCysLeuAspTyrMetGlnGlnAsnHisIlePheVal 320
QY 2688 TTTTCACTCAGCTGGCCAAACAAGGGGCGGAATCAGTTTAAAGCGTCAATTTCAAACT 2747
Db 321 PheSerThrGlnLeuIaAsnLysGlyAlaGluSerValLeuSerGlyGlnPheGlnThr 340
QY 2748 ATTATTGCGTATCACTGCACTCAGCCTGCTACAAAAGCTTCTGGAAGACTTTTATG 2807
Db 341 IleIleAlaTyrHisCysThrGlnProAlaThrLysSerPheLeuGluAspPhePheMet 360
QY 2808 AAAAAACCTTTAAAGATTAAACAAGTTACAGCGGCACAAATTCGTCGATGCGATGAGTA 2867
Db 361 LysAsnProLeuLysIleAsnLysLeuGlnArgHisAsnSerValGlyMetProIle 380
QY 2868 GGCATGGGCGAGTTGSACTAACTCCTCCTAATCCTGTAGCCAAAATAACACAACAGCAG 2927
Db 381 GlyMetGlyGlnValGlyLeuThrProProAsnProValAlaLysIleThrGlnGln 400
QY 2928 CCACATACAAGAAGCCGTAGCCCTATTGAAACCCCAATTCAATCAACATGAAGAAACAGCAA 2987
Db 401 ProHisThrLysThrValGlyLeuLeuLysProGlnPheAsnGlnHisGluAsnSerLys 420
QY 2988 CGTAGTACTGTAAAGCGCGCTAGCAACTCTTTTGTCAACCAAGTCTGATCCTATGGCAAC 3047
Db 421 ArgSerThrValSerAlaProSerAsnSerPheValAspGlnSerAspPrometGlyAsn 440
QY 3048 GAAACTGAATTGATGTGCTGGGAAGGCGGATCCTCAACACCAAGTAGTCTGGACAAAAC 3107
Db 441 GluThrGluLeuMetCysTrpGluGlyGlySerSerAsnThrSerArgSerGlyGlnAsn 460
QY 3108 TCACGAAATCATGTAGACAGTATCAGTACATCCAGCAGTCAAGGCAATAAAGATACTG 3167
Db 461 SerArgAsnHisValAspSerIleSerThrSerSerGluSerGlnAlaIleLysIleLeu 480
QY 3168 GAAAGCAGCTGCGTTGATTGGGACAGGTACAAAAGAAAGCAGTCCGTCCTGCAACT 3227
Db 481 GluAlaIaGlyValAspLeuGlyGlnValThrLysGlySerAspProGlyLeuThrThr 500
QY 3228 GAAAAACAACATTGTATCACTGCAAGAGAGTTAAGTTCCAGAGAAACCTTACACACAA 3287

Db 501 GluAsnAsnIleValSerLeuGlnGlyValLysValProAspGluAsnLeuThrProGln 520
QY 3288 CAGCGCAACATCGGAAGAACAAGTTGGCAAAAATAAAAAATGAATCAATTTCTTTT 3347
Db 521 GlnArgGlnHisArgGluGluGlnGlnLeuAlaLysIleLysLysMetAsnGlnPheLeuPhe 540
QY 3348 CCTGAAAATGAGAAATTCAGTAGAGACTAAATGTAAGCTCACAGATAACAATAATCCAGGA 3407
Db 541 ProGluAsnGluAsnSerValGlyAlaAsnValSerSerGlnIleThrLysIleProGly 560
QY 3408 GATTTAATGATGGGATGTCCGGGTGCGGAGCGGATCTATTATTAATCCGACGATGCGA 3467
Db 561 AspLeuMetMetGlyMetSerSerGlyGlyGlyGlySerIleIleAsnProThrMetArg 580
QY 3468 CAACTGCATATGCCAGGTAAACGCCCAATCCGAGCTCTTATCCGCGACAAGTTCAGACTT 3527
Db 581 GlnLeuHisMetProGlyAsnAlaLysSerGlnLeuLeuSerAlaThrSerSerGlyLeu 600
QY 3528 TCGGAAGATGTATGATCCATCCAGGGAGTGTATATCAGATATGGTGCCGTAAATAGATGT 3587
Db 601 SerGluAspValMetHisProGlyAspValIleSerAspMetGlyAlaValIleGlyCys 620
QY 3588 AATAATAATCAAAAAACCAAGTGTGCAATGTGATCTGAGTAGAGTGTGTCACTGGAACA 3647
Db 621 AsnAsnAsnGlnLysThrSerValGlnCysGlySerGlyValGlyValValThrGlyThr 640
QY 3648 ACTGACGTGAGTAATGTCAATATGATGCTCAAGCTCCGGCGCGCCGCAATGCGCAAT 3707
Db 641 ThraIaAlaGlyValAsnValAsnMetHisCysSerSerSerSerGlyAlaProAsnGlyAsn 660
QY 3708 ATGATGGGAAGCTTACCGGATATGCTAGCCTCGTTTGGCAACACAGCTGCAACGTATC 3767
Db 661 MetMetGlySerSerThrAspMetLeuAlaSerPheGlyAsnThrSerCysAsnValIle 680
QY 3768 GGAACGGCCCCCAGATATGTTAAAGAGTTTAAATCAAGATAGCCGAACCAATTCACAT 3827
Db 681 GlyThraIaProAspMetSerLysGluValLeuAsnGlnAspSerArgThrHisSerHis 700
QY 3828 CAAGGGGAGTTGCTCAAAATGAGTGTGAGATTCGAAGATTCAAATCAATTTTGAAGAACGC 3887
Db 701 GlnGlyGlyValAlaGlnMetGluTrpSerLysIleGlnHisGlnPhePheGluArg 720
QY 3888 CTCAAGGGGGGCAAGCCCAAGACAAGTCACTGGAACCTGTAGTACCACAACAGCAAAACCT 3947
Db 721 LeuLysGlyGlyLysProArgGlnValThrGlyThrValValProGlnGlnGlnThrPro 740
QY 3948 TCTGATCTGTGTGAAACTCGTTAAACAACACAGTCCGACCCCTGCAAGTCCACTCCT 4007
Db 741 SerGlySerGlyGlyAsnSerLeuAsnAsnGlnValArgProLeuGlnGlyProProPro 760
QY 4008 CCTTACCACCTCCATCCAGAGATCTGCGTCAAGTCAACCAATAGCCACTCAATCGCCAATCCC 4067
Db 761 ProTyrHisSerIleGlnArgSerAlaSerValProIleAlaThrGlnSerProAsnPro 780
QY 4068 TCGAGTCCAAACAATCTATCTCTCCCGTCAAGCGGACAACCGCAGCAGTCAAGGATTG 4127
Db 781 SerSerProAsnAsnLeuSerLeuProSerProArgThrThraIaAlaValMetGlyLeu 800
QY 4128 CCGACCAACTCTCTAGCATGAGTGAACAGATCAATTATCTGATCTGTTCCGCAAGCT 4187
Db 801 ProThrAsnSerProSerMetAspGlyThrGlySerLeuSerGlySerValProGlnAla 820
QY 4188 AATACTTGACGTTCAAGCAGGACACAACAAGTGTCTCAGCAAAACAAGACTGTTTTT 4247
Db 821 AsnThrSerThrValGlnAlaGlyThrThrThrValLeuSerAlaAsnLysAsnCysPhe 840
QY 4248 CAGGACAGACCCCATCGCCGTCAATCAAAATCGTAGTAAGAATACCGGATCGTCAAGC 4307
Db 841 GlnAlaAspThrProSerProSerAsnGlnAsnArgSerArgAsnThrGlySerSerSer 860
QY 4308 GTTCTTACGCATTAATTAAAGCAGCAACCCCAAGTACCCCTTATCTCATCTTATCCCAAG 4367
Db 861 ValLeuThrHisAsnLeuSerSerAsnProSerThrProLeuSerHisLeuSerProLys 880

| | | | |
|----|------|--|------|
| QY | 4368 | GAATTGAGTCTTTCGGTCAGTCTCTGCTGCTGATGTTAAATTTGTTAAATTTTAA | 4427 |
| Db | 881 | GluphegIuserPheGlyGlnSerSer----- | 889 |
| QY | 4428 | GACAAATCAAAATATGAATTGCGTTAATAATGAATTATATATTACATACTCGAAATTTG | 4487 |
| Db | 889 | ----- | 889 |
| QY | 4488 | ATAGAAAAAATCAGGAATAGAAAAAATAATTAATTTCCGGACCGCCATCCATTTCTTG | 4547 |
| Db | 889 | ----- | 889 |
| QY | 4548 | AATCCAATTTCTGAGTGAATTGTTAGAGATAATCTACTAATTAATAATTAACACGAAAT | 4607 |
| Db | 889 | ----- | 889 |
| QY | 4608 | CATATCCGTTAATTGAAAACTATGTTTAAATAAGAAATTAATAATATGTTAATTATA | 4667 |
| Db | 889 | ----- | 889 |
| QY | 4668 | ATATTTCTACAGTGTATACATGAAGAAAGTAGCGACCAAGCCACAGGTACGGTCA | 4727 |
| Db | 890 | -----Ala--GlyaspheMetIyserArgArgProSerProGlnGlyGlnArgSerP | 907 |
| QY | 4728 | CAGTAATAGTCTAATAGAGCGCAATAAAGATGTACGATTTGCTGCATCCAGTCTCTGTT | 4787 |
| Db | 907 | roValasnSerLeuIleGluIlaasnlysbaspValArgPheAlaIaSerSerProGlyP | 927 |
| QY | 4788 | TTAACCCCGCATCCACATATGCAAGCAATTCAAATTTACGATTAAACGCTATATAATGG | 4847 |
| Db | 927 | heasnProHisProHisMetGlnSerAsnSerAsnSerAlaIeuAsnAlaTyrIysMetG | 947 |
| QY | 4848 | GCTCTACCAATATACAGATGAGGTAAATATTTAAATAATTTAATTAACGTTTGTGT | 4907 |
| Db | 947 | IyserThrAsnIleGlnMetGlu--Arg----- | 955 |
| QY | 4908 | AATTTATCTTTTTCAGCGTCAAGCATCAGCGCAAGGTGATCCGTACAATTTAGTCG | 4967 |
| Db | 956 | -----GlnAlaSerAlaGlnGlyGlySerValGlnPheSerAr | 968 |
| QY | 4968 | GCGCTCCGATAATATTTCCGCTAAATCCCAATAGTGCAATCGGCCGCCACCAACAAGAT | 5027 |
| Db | 968 | gArgSeraspAsnIleProIeuasnProAsnSerGlyAsnArgProProProAsnlySme | 988 |
| QY | 5028 | GACCCAAAACCTTGCATCCAATCTCTTCTTTGGCACAAATGTCCCAACACTTAACAAGTTG | 5087 |
| Db | 988 | tThrGlnasnPheaspProIleSerSerIeuAlaGlnMetSerGlnGlnLeuThrSerCy | 1008 |
| QY | 5088 | CGTGTCCAGCATGGGTAGTCCAGCCGGAACCTGGTGTATGACGATGAGGGGGTCCGGG | 5147 |
| Db | 1008 | sValSerSerMetGlySerProAlaGlyThrGlyGlyMetThrMetMetGlyGlyProGl | 1028 |
| QY | 5148 | ACCGTCCGACATCAATATTTGACCATGGAATAATTTCCGGACTAGATGGATCAGGAATAGA | 5207 |
| Db | 1028 | yProSeraspIleasnIleGluHisGlyIleIleSerGlyIleuaspGlySerGlyIleas | 1048 |
| QY | 5208 | TACCATTAATCAAAATACTGTTCATTTCAATGAATGTCTGAATGAACCTCAATGGGTCCCG | 5267 |
| Db | 1048 | pThrIleasnGlnasnAsnCyshisSerMetAsnValValMetAsnSerMetGlyProAr | 1068 |
| QY | 5268 | AATGCTGAATCTTAAATGTGCGTAGCAGCGCGGTCCAAATGGAACCGCTGCTTAAATCC | 5327 |
| Db | 1068 | gMetIeuasnProIysMetCysValAlaGlyGlyProAsnGlyProProGlyPheasnPr | 1088 |
| QY | 5328 | TAATTTCCCAATGTTGATTAAGAGAATTCATAGGGTCTGGCTGTGCTCAGCAAA | 5387 |
| Db | 1088 | oasnSerProAsnGlyGlyIleuArgGluAsnSerIleGlySerGlyCybsGlySerAlaAs | 1108 |
| QY | 5388 | CTCTTCAAACCTTCAAGGGGTGTTCCACCTGTGTCAGAAATGATGGGTGGAATGCCAGT | 5447 |
| Db | 1108 | nSerSerasnPheGlnGlyValValProProGlyAlaArgMetMetGlyArgMetProVa | 1128 |

| | | | |
|----------|-----------|---|------|
| QY | 5448 | CAATTTGGTTGCAATTTCATCCGAATATTCAGSTAAAGCGGAGTACCCCAACACCAT | 5507 |
| Db | 1128 | IAAnpHeGlySerAsnPhneAsnProAsnIleGlnValIysValAserThrProAsnThrI | 1148 |
| QY | 5508 | ACAATACATGCCAGTAAGGGCACAGAACGCCAACAAACAATAACAACAATGAGCTAATAA | 5567 |
| Db | 1148 | eGlnTyMetProValArgAlaGlnAsnAlaAsnAlaAsnAlaAsnAlaAsnAlaAsnAs | 1168 |
| QY | 5568 | TGTGCGAATGCCACCTAGTCTGGAATTTTTCAGAGGTACGGTAAACCTCAATGGGTGC | 5627 |
| Db | 1168 | nValArgMetProProSerLeuGluPheLeuGlnArgTyAlaAsnProGlnMetGlyAl | 1188 |
| QY | 5628 | TGTAGGCAATGGGTGCGCAATATATGCCACCATCAGCCAGCGAGTACTCTGGAATGCC | 5687 |
| Db | 1188 | aValGlyAsnGlySerProIleCysProProSerAlaSerAspGlyThrProGlyMetPr | 1208 |
| QY | 5688 | AGGATTGATGGCGGACCAAGAGCCGGAGGTATGCTAATGAATTCTTCCGAGACCAACA | 5747 |
| Db | 1208 | oGlyLeuMetAlaGlyProGlyAlaGlyGlyMetLeuMetAsnSerSerGlyGlnGlnH | 1228 |
| QY | 5748 | CCAGAACCAAGATCACAAACAATCCTGGGGCAAGCAATGGTATTAACTTCTTTCAGAA | 5807 |
| Db | 1228 | sGlnAsnLysIleThrAsnAsnProGlyAlaSerAsnGlyIleAsnPheGlnAsnCy | 1248 |
| QY | 5808 | CAATCAATGTCTATTGTGTGACGAGAGGGGTGATTACCCGGCCATGACGGATCAATGAA | 5867 |
| Db | 1248 | sAsnGlnMetSerIleValAspGluGluGlyGlyLeuProGlyHIsAspGlySerMetAs | 1268 |
| QY | 5868 | TATTGGTCAACCATCTATGATTAAGGGGCGATGGCGTCCACATGGCCATGGGCCAAATG | 5927 |
| Db | 1268 | nIleGlyGlnProSerMetIleArgGlyMetArgProHIsAlaMetArgProAsnValMe | 1288 |
| QY | 5928 | GGGTGCGCGGATGCCACCCCGTTAACAGGCAAAATTCAGTTTGCACAGTATCGGATGG | 5987 |
| Db | 1288 | tGlyAlaArgMetProProValAsnArgGlnIleGlnPheAlaGlnSerSerAspGlyI | 1308 |
| QY | 5988 | TGACTGTGTGCGGGATCCGTCATCATTTTTCACCTAACGCTTCTGCAACAGCGCTGGA | 6047 |
| Db | 1308 | eAspCysValGlyAspProSerSerPhePheThrAsnAlaSerCysAsnSerAlaGlyPr | 1328 |
| QY | 6048 | ACACATGTTTGGATCAGCAACAACAGGCCAATCAGCCTTAAGACACACAACATAAAGAC | 6107 |
| Db | 1328 | oHisMetPheGlySerAlaGlnGlnAlaAsnGlnProLysThrGlnHIsIleLysAsnI | 1348 |
| QY | 6108 | ACCTAGTGAATGTGTCAAAAACAATCGGGACTTGACGTGGCACAAGGCGAGATCCA | 6167 |
| Db | 1348 | eProSerGlyMetCysGlnAsnGlnSerGlyLeuAlaValAlaGlnGlyGlnIleGlnLe | 1368 |
| QY | 6168 | GCATGGGCAAGGACATGGCGCAGGGTCAGTCTTTAATTGGACCTACTAATAATATTAA | 6227 |
| Db | 1368 | uHisGlyGlnGlyHIsAlaGlnGlyGlnSerLeuIleGlyProThrAsnAsnLeuMe | 1388 |
| QY | 6228 | GTCAACTGCCGGAAGTGTCAAGTCTACTAACGGTGTCTCTGGCATCAATTTCTGAGTCC | 6287 |
| Db | 1388 | tSerThrAlaGlySerValSerAlaThrAsnGlyValSerGlyIleAsnPheValGlyPr | 1408 |
| QY | 6288 | CTCTTCTACGGACCTGAAGTATGCCCGACAATATCATAGTTTTCAGCAGCAGTTATATGC | 6347 |
| Db | 1408 | oSerSerThrAspLeuLysTyAlaGlnGlnTyHIsSerPheGlnGlnLeuTyAl | 1428 |
| QY | 6348 | TACCAACACACCAGAAGTCAACAACAACAGCATATGACACAGCAGCAGAGCAACATGAT | 6407 |
| Db | 1428 | aThrAsnThrArgSerGlnGlnGlnGlnHIsMetHIsGlnGlnHIsGlnSerAsnMetI | 1448 |
| QY | 6408 | AACAATGCCGCGCAATTATACCAACAATCCAAGCTTCTTGTCAACAADA | 6456 |
| Db | 1448 | eThrMetProProAsnLeuSerProAsnProThrPhePheValAsnLys | 1464 |
| RESULT 2 | | | |
| ID | ABW01527 | | |
| XX | ABW01527 | standard; protein; 1464 AA. | |
| AC | ABW01527; | | |

XX 15-JAN-2004 (first entry)
DT Drosophila species legless (lgs) protein.
DE
XX
KM Legless protein; lgs, cell fate disorder; blood disease; gene therapy;
KW cancer; tissue regeneration; tissue repair; cyostatic.
XX
OS Drosophila sp.
XX
PN US2003114413-A1.
XX
PD 19-JUN-2003.
XX
PF 19-DEC-2002; 2002US-00322579.
XX
PR 28-JUL-2000; 2000US-0221502P.
PR 27-JUL-2001; 2001US-00915543.
XX
PA (UYZU-) UNIV ZURICH.
XX
PI Basler K, Brunner E, Froesch B, Kramps T, Peter O;
XX
DR WPI; 2003-829432/77.
DR N-PSDB; AAD62641.
XX
XX Novel lgs polypeptide useful for isolation of lgs-binding proteins,
PT diagnosing disorders of cell fate, treating diseases such as cancer.
XX
PS Claim 5; Fig 2; Opp; English.
XX
XX The invention relates to novel legless (lgs) proteins and polynucleotides
CC encoding such proteins. lgs sequences are useful for the treatment of
CC disorders of cell fate such as differentiation or proliferation. The
CC invention is used to treat blood disease or a cancerous condition
CC characterised by over-stimulation of the Wnt pathway such as colon,
CC breast, head and neck, brain, thyroid, medulloblastoma or skin cancer and
CC is administered to prevent progression from a pre-neoplastic or non-
CC malignant condition to a neoplastic or malignant state. It is
CC administered to promote tissue regeneration and repair. The invention is
CC also useful in the therapy of diseases cost by an over-activation of Wg
CC pathway. It is useful for reducing lgs gene expression in an invertebrate
CC or vertebrate organism or an invertebrate or vertebrate cell line. The
CC invention is also useful in gene therapy. The present sequence is
CC Drosophila species legless (lgs) protein
XX
SQ Sequence 1464 AA;

Alignment Scores:
Pred. No.: 0 Length: 1464
Score: 7257.00 Matches: 1460
Percent Similarity: 73.01% Conservative: 1
Best Local Similarity: 72.96% Mismatches: 3
Query Match: 60.29% Indels: 537
DB: 7 Gaps: 5

US-10-664-859-1 (1-6909) x ABW01527 (1-1464)

QY 468 ATGCCCGCAGTCCACCAACACGCGCAACCAACTCCGATGCTCTCTCAACAAGT 527
DB 1 MetProArgSerProThrGlnGlnGlnProGlnProAsnSerAspAlaSerSerThrSer 20
QY 528 GCATCTGATCAATCCTTGAGCAGCGATCGAATGGGACTCGCGCGGAGAGAAGT 587
DB 21 AlaSerGlySerAsnProGlyAlaAlaIleGlyAsnGlyAspSerAlaAlaSerArgSer 40
QY 588 TCTCCGAGAGACCTTAATAGCGAACCTTTTCTACTTGTGCGCGGTAAGACTTGTATT 647
DB 41 SerProLysThrLeuAsnSerGluProPheSerThrLeu-Ser----- 54
QY 648 GATTTCCTTTGTGCGGAATTATACAACTTTGTGTGTTCCAGATCAATTAATTGAC 707
DB 55 -----ProAspGlnIleLysLeuTh 61

QY 708 GCCAGAGAGGCACTGAGAAAAGCGGACTATCACTAGTATTAAGCTGCCACTGGAGG 767
DB 61 rProGluGlyThrGluLysSerGlyLeuSerThrSerAspLysAlaIleThrGlyGly 81
QY 768 AGCCCCAGGCGAGTGAATTAATCTGCCCGGAGGACAAACTATGCTAAGCGAGAAGTCTAC 827
DB 81 yAlaProGlySerGlyAsnAsnLeuProGluGlyGlnThrMetLeuArgGlnAsnSerTh 101
QY 828 GAGCACAATCACTCGTGCCTAGTGCCTTCTCCACAAAACCTCCAGTGAAGTGAATAG 887
DB 101 rSerThrIleAsnSerCysLeuValAlaSerProGlnAsnSerSerGluHisSerAsnSe 121
QY 888 CAGCAATGTGTCTGCTACAGTGGCGCTTACTCAGATGTAGATTGTGACGACAATCGAA 947
DB 121 rSerAsnValSerAlaThrValGlyLeuThrGlnMetValAspCysAspGluGlnSerLy 141
QY 948 GAAAAACCAATGTAGTGTGAAGACGAGGAAGCTGTGAAGACTGCCCTACAAATGCTTTA 1007
DB 141 sLysAsnLysCysSerValLysAspGluGluAla----- 152
QY 1008 AAATTTAAATGTATTGGCGTTCACCTTTGTTAATCATTTAATTGTTTTTTTGCTA 1067
DB 152 ----- 152
QY 1068 TACTTACAATTTTACTTTTAAACCTTGTAACCTGACTAAACCTCGGAGCTCGATCAA 1127
DB 152 ----- 152
QY 1128 AACAGACATTTTCTTGGAACCGTAATTAAGCTCATAAAAATATTAAATTCATTGATGA 1187
DB 152 ----- 152
QY 1188 ATGCATATCATAGATGTACTCAAAACATCTCAGAAAGACCTCAAAATTGATCAACTAATT 1247
DB 152 ----- 152
QY 1248 AGTTGAGAAAAAATTGCTGTACTTTTAAGAAATATATTAAATTTGCTGAGTGA 1307
DB 152 ----- 152
QY 1308 AATGATATAATAGTCACAATAATTTTAACTGTAAGACATTTGAATAGCCGTG 1367
DB 152 ----- 152
QY 1368 CTACGAGATGTACTAGACGCGGTGAAGCTAATTTTATTAAAGCTGTCTAAT 1427
DB 152 ----- 152
QY 1428 ATTCCATAACCAATTAATGTCCCATTTCAGAATAAGTTCTAATAAGCAAAAGTCAAGC 1487
DB 153 -----GluIleSerSerAsnLysAlaLysGlyGlnAl 163
QY 1488 AGCTGTGGCGGCTGCGAAACAGGTTCTACATCCAGTTTGACTGTCAAGAGAAGACCCAC 1547
DB 163 aAlaGlyGlyGlyCysGluThrGlySerThrSerSerLeuThrValLysGluGluProth 183
QY 1548 CGATGTCTTAGGCAATTAGTAATATGAAAAAGAAAGAAAGAAATCATTCGCCAAC 1607
DB 183 rAspValLeuGlySerLeuValAsnMetCysLysGluGluArgGluAsnHisSerProth 203
QY 1608 GATGTCCCTGTGTTGTTTGTTCATTTGGTATGTCACAGAGCAACTCGCTACACCGGG 1667
DB 203 rMetSerProValGlyPheGlySerIleGlyAsnAlaGlnAspAsnSerAlaThr----- 221
QY 1668 TAAGTTTAAGAGATCCATATTAAGCAATAACAAGAAATTAATGTCAGTTACCAATTTTA 1727
DB 221 ----- 221
QY 1728 TTTGATAGTCAAGAAGAACTACTATAGCGAATATCTCCTGCTTTAATTTAATTAG 1787
DB 221 ----- 221

QY 1788 GAAATACGAATATTCTTAATTTGTAAAAATAAAATTGATTAACTAGAAATTTAAAAAC 1847
Db 221 ----- 221
QY 1848 CTTTGAATTAGACATACCCTTCCAAAAATCAGTAATCATTGGGAACGAGAGTGTGTC 1907
Db 221 ----- 221
QY 1908 CCGAAGAGACTACTATAAAACCTTTGAGCTATCTGATACTGACGCTACTAAAAATGA 1967
Db 221 ----- 221
QY 1968 TTAGTTAGGAAAAATGGGTGTAATTTTGTAGGAAGTTTTCATTTTAGAGAAGAAATGTGATT 2027
Db 221 ----- 221
QY 2028 ATTTTATTAACCCCTTCAAGCGGAACTACATTGTTCTACGATATTTTGGAAAAACAAC 2087
Db 221 ----- 221
QY 2088 TGGTTAAGTTGAAAAGTGCCTATTAACAGAAATCCACGGTTTCAAACTAACCAGGTT 2147
Db 221 ----- 221
QY 2148 TTTGATTTAATTTTGATTAAATGAGAAATTATCACACTTCAGTTAAATGTTTAATTGCA 2207
Db 221 ----- 221
QY 2208 TTAAGTCGGAACAATCACAGAGATTTTCATTTTGCGTATATATAGAAAGTCGCTTC 2267
Db 221 ----- 221
QY 2268 ACACTCTTGGCGCGCTTCACCACTACGTGAGATTCCGCCCGCAGTATTATATAGAT 2327
Db 221 ----- 221
QY 2328 GATTTCAGAGTTATTTAATTTTATGGTGTAATTTTAATAATATCTTAATTATCATTT 2387
Db 221 ----- 221
QY 2388 TACATAGTTAAATTGAAAGAATTTCAACGACAGTACCACGGAAAAAAGCATCGTCC 2447
Db 222 -Pro-ValLysIleGlyValGlyAlaGlyThrGlyAsnLeuLeuThrAlaAsnAlaAsnGly 240
QY 2448 TTGACAATGAATTAATGACGAATGAGCATGGAAGGCTGCAATCAGTTGAATCCCGATTTT 2507
Db 241 LeuThrMetAsnAsnAspGluMetSerMetGluGlyCysAsnGlnLeuAsnProAspPhe 260
QY 2508 ATCAATGAATCTTTAAATAATCCTGCAATTTGAGCATATTAGTAAGCGGAGTAGACCA 2567
Db 261 IleAsnGlnSerLeuAsnAsnProAlaIleSerSerIleLeuValSerGlyValGlyPro 280
QY 2568 ATACCCGGAATCGAGTTGGAGCGGGGAGCGGAATTTTATTGACTGCCAACGCCAATGGA 2627
Db 281 IleProGlyIleGlyValGlyAlaGlyThrGlyAsnLeuLeuThrAlaAsnAlaAsnGly 300
QY 2628 ATCTCCTCGGGTAGCAGTAATTTGTTGGATTACATGCAACAGCAAAATCACAATATTCGTG 2687
Db 301 IleSerSerGlySerSerAsnGlyLeuAspTyrMetGlnGlnGlnAsnHisIlePheVal 320
QY 2688 TTTTCAACTCAGCTGGCCCAACAAGGGGGCGGAATCAGTTTAAAGCGGTCAATTTCAAACT 2747
Db 321 PheSerThrGlnLeuAlaAsnLysGlyAlaGlnSerValLeuSerGlyGlnPheGlnThr 340
QY 2748 ATTATTGCGTACACTGCACTCAGCCTGCTACAAAAAGCTTCTGAGAGACTTTTATG 2807
Db 341 IleIleAlaIleTyrHisCysThrGlnProAlaThrLysSerPheLeuGluAspPheMet 360
QY 2808 AAAAACCCTTTAAAGATTAAACAAGTTACAGCGGCACAATTCGTCGGTATGCCATGATA 2867
Db 361 LysAsnProLeuLysIleAsnLysLeuGlnArgHisAsnSerValGlyMetProTyrIle 380
QY 2868 GGCATGGGGCAGGTGGACTAACTCCTCTTAATCTGTAGCCAAAAATTAACACACAGCAG 2927

Db 381 GlyMetGlyGlnValGlyLeuThrProProAsnProValAlaLysIleThrGlnGlnGln 400
QY 2928 CCAGATACAAGACCCTAGCCCTATTGAAACCCCAATTCAATCAACATGAAAAACAGCAAA 2987
Db 401 ProHisThrLysThrValGlyLeuLeuLysProGlnPheAsnGlnHisGluAsnSerLys 420
QY 2988 CGTAGTACTGTAAAGCGCGCCTAGCAACTCTTTTGTGACCAAGTCTGATCTATGGGCAAC 3047
Db 421 ArgSerThrValSerAlaProSerAsnSerPheValAspGlnSerAspPrometGlyAsn 440
QY 3048 GAAACTGAATTGATGTGCTGGGAAGCGCGATCCTCAAAACACCAAGTAGTCTGGACAAAC 3107
Db 441 GluThrGluLeuMetCysTrpGluGlyGlySerSerAsnThrSerArgSerGlyGlnAsn 460
QY 3108 TCACGAAATCATGTAGACAGTATCATCATCCAGCGAGTTCACAGGCAATAAAGATACTG 3167
Db 461 SerArgAsnHisValAspSerIleSerThrSerSerGluSerGlnAlaIleLysIleLeu 480
QY 3168 GAAGCAGCTGCGGTTGATTGGGACAGGTACAAAAAGAAAGCATCCTGCGCTGACAACT 3227
Db 481 GluAlaAlaGlyValAspLeuGlyGlnValThrLysGlySerAspProGlyLeuThrThr 500
QY 3228 GAAACAACATTTGATATCACTGCAAGAGTTAAGTTCCAGACGAAACCTTACACCACAA 3287
Db 501 GluAsnAsnIleValSerLeuGlnGlyValLysValProAspGluAsnLeuThrProGln 520
QY 3288 CAGCGGCAACATCGGGAAGACAGCTTGGCAAAAAATAAAAAATGAATCAATTTCTTTT 3347
Db 521 GlnArgGlnHisArgGluGluGlnLeuAlaLysIleLysLysMetAsnGlnPheLeuPhe 540
QY 3348 CCTGAAAAATGAGAATTACAGTAGAGCTAATGTAAAGTCCACAGATAACAAAAATTCAGCA 3407
Db 541 ProGluAsnGluAsnSerValGlyAlaAsnValSerSerGlnIleThrLysIleProGly 560
QY 3408 GATTTAATGATGGGATGTCGGGTGGCGGAGCGGAGTCTATTTAAATCCGACGATGCGA 3467
Db 561 AspLeuMetMetGlyMetSerGlyGlyGlyGlySerIleIleAsnProThrMetArg 580
QY 3468 CAACTGCATATGCCAGTAACGCCAAATCGGAGCTCTTATCGCGGACAAAGTTCAGACACT 3527
Db 581 GlnLeuHisMetCProGlyAsnAlaLysSerGluLeuLeuSerAlaThrSerSerGlyLeu 600
QY 3528 TCGGAAGATGTAATGCATCCAGGGAGTGTATATCAGATATGGGTGCCGTATAGAGATGT 3587
Db 601 SerGluAspValMetHisProGlyAspValIleSerAspMetClyAlaValIleGlyCys 620
QY 3588 AATATAATCAAAAAACCAAGTGTGCAATGTGGATCTGAGTAGGTGTGTCACTGGAACA 3647
Db 621 AsnAsnAsnGlnLysThrSerValGlnCysGlySerGlyValGlyValIleThrGlyThr 640
QY 3648 ACTGACCTGAGTAATATGTCATATGCAATGTCTCAAGCTCCGGCGCCCGGAATGGCAAT 3707
Db 641 ThrAlaAlaGlyValAsnValAsnMetHisCysSerSerSerGlyAlaProAsnGlyAsn 660
QY 3708 ATGATGGGAAGCTCTACGATATGCTAGCCTCGTTTGGCAACAAGCTGCAACGTCAATC 3767
Db 661 MetMetClySerSerThrAspMetLeuAlaSerPheGlyAsnThrSerCysAsnValIle 680
QY 3768 GGAAGCGCCCAAGATATGTCTAAGGAAGTTTAAATCAAGATAGCCGAACCCATTCACAT 3827
Db 681 GlyThrAlaProAspMetSerLysGluValLeuAsnGlnAspSerArgThrHisSerHis 700
QY 3828 CAAGGGGAGTTGCTCAAAATGAGTGTGGAAGATTCAACATCAATTTTTCGAGAAGACGC 3887
Db 701 GlnGlyGlyValAlaGlnMetGluTrpSerLysIleGlnHisGlnPhePheGluGluArg 720
QY 3888 CTCAAGGGGGCAAGCCCAAGACAAGTCACTGGAATGTAGTACCAACAAGCAAAACCCCT 3947
Db 721 LeuLysGlyGlyLysProArgGlnValThrGlyThrValValProGlnGlnGlnThrPro 740
QY 3948 TCTGATCTGTGGAACCTCGTTAAACAACCAAGGTGCGACCCCTGCAAGGTCCACCTCCT 4007

Db 741 SerGlySerGlyGlyAsnSerLeuAsnAsnGlnValArgProLeuGlnGlyProProPro 760
QY 4008 CCTTACCACCTCCATCCAGAGATCTGCGTCAAGCAATAGCCACTCAATCGCCCAATCCC 4067
Db 761 ProTyrHisSerIleGlnArgSerAlaSerValProIleAlaThrGlnSerProAsnPro 780
QY 4068 TCGAGTCCAAACAATCTATCTCTCCCGTCAACCGCGGACACACCGCAGCATGCGATTG 4127
Db 781 SerSerProAsnAsnLeuSerLeuProSerProArgThrThrAlaAlaValMetGlyLeu 800
QY 4128 CCGACCAACTCTCTAGCATGATGGAACAGATCATTTATCTGGATCTGTTCCGCAAGCT 4187
Db 801 ProThrAsnSerProSerMetAspGlyThrGlySerLeuSerGlySerValProGlnAla 820
QY 4188 AATACTTCGACGGTTCAGGCAGGCACAACAAGTCTCTCAGCAACAACAAGACTGTTT 4247
Db 821 AsnThrSerThrValGlnAlaGlyThrThrThrValLeuSerAlaAsnLysAsnCysPhe 840
QY 4248 CAGCAGACACCCCATCGCCGTCAAAATCAAAATCGTAGTAAGAATACCGGATCGTCAAGC 4307
Db 841 GlnAlaAspThrProSerProSerAsnGlnAsnArgSerArgAsnThrGlySerSerSer 860
QY 4308 GTTCTTACGCATTAATTAAAGCAGCAACCCCAAGTACCCCTTATCTCATCTATCCCAAG 4367
Db 861 ValLeuThrHisAsnLeuSerSerAsnProSerThrProLeuSerHisLeuSerProLys 880
QY 4368 GAATTGTAGTCTTTCGGTCAGTCCCTCTGCTGGTATGTTATATTGTTTAATTTTTTAA 4427
Db 881 GluPheGluSerPheGlyGlnSerSer----- 889
QY 4428 GACAAATCAAAATATGAATTGCGTTAATTAAGTTATATATTACATACTCGAAATTGG 4487
Db 889 ----- 889
QY 4488 ATAGAAAAAATCAGGAATAGAAAAAATAAATATTTCGCGACCGCCCATTTCTTG 4547
Db 889 ----- 889
QY 4548 AATCCAATTTCGGAGTGATTGTTAGAGATACTACTATTAAATTAACAAGAAATT 4607
Db 889 ----- 889
QY 4608 CATATCCGTTAATTGAAAAATCACTATTGTTTAATAAGAAATTAAAAATATGTTATTATA 4667
Db 889 ----- 889
QY 4668 ATATTCTACAGGTGATTAACATGAAAAAGTAGCGCAACCAAGCCACAGGGTCAAGGTCAC 4727
Db 890 -----Ala--GlyAspAsnMetLysSerArgArgProSerProGlnGlyGlnArgSerP 907
QY 4728 CAGTAATAGTCTAATAGAGGCAAAATAAGATGACGATTGCTGCATCCAGTCTGTT 4787
Db 907 roValAsnSerLeuIleGlnAlaAsnLysAspValArgPheAlaAlaSerSerProGlyP 927
QY 4788 TTAACCCGCATCCACATATGCAAAAGCAATTCAAAATTCAGCATTTAAACGCTATAAAATGG 4847
Db 927 heAsnProHisProHisMetGlnSerAsnSerAsnSerAlaLeuAsnAlaTyrLysMetG 947
QY 4848 GCTCTACCAATATACAGATGAGGTAATATATTAAATATTATTAAAGTTTGTGTT 4907
Db 947 LysSerThrAsnIleGlnMetGlu--Arg----- 955
QY 4908 AATTATCTTTTTCAGCGTCAAGCATCAGCGCAAGTGATCCGTTACAATTTAGTCG 4967
Db 956 -----GlnAlaSerAlaGlnGlyGlySerValGlnPheSerAr 968
QY 4968 GCGCTCCGATATATTCCGCTAAATCCCAATAGTGGCAATCGCGCCACCAACAAGAT 5027
Db 968 gArgSerAspAsnIleProLeuAsnProAsnSerGlyAsnArgProProAsnLysMe 988
QY 5028 GACCCAAAACCTTCGATCCAATCTCTTTGGGCAAAATGTCCCAACAACCTAACAGTTG 5087
Db 988 tThrGlnAsnPheAspProIleSerSerLeuAlaGlnMetSerGlnGlnLeuThrSerCy 1008

QY 5088 CGTGTCCAGATGGGTAGTCCAGCCCGGAACCTGTGTATGACGATGAGGGGCTCCGGG 5147
Db 1008 sValSerSerMetGlySerProAlaGlyThrGlyGlyMetThrMetGlyGlyProG 1028
QY 5148 ACCGTCCGACATCAATATTAGCATGGAATAATTTCCGGACTAGATGATCAGGAATAGA 5207
Db 1028 yProSerAspIleAsnIleGlnIleGlyIleIleSerGlyLeuAspGlySerGlyIleAs 1048
QY 5208 TACCATAAATCAAAATACTGTCAATGAATGTCGTAATGAACCTAATGGTCCCGG 5267
Db 1048 pThrIleAsnGlnAsnAsnCysHisSerMetAsnValValMetAsnSerMetGlyProAr 1068
QY 5268 AATGCTGAATCCTAAATGTGCGTAGCAGCGGTCCTCAATGGAACCGCTGCTTAATCC 5327
Db 1068 gMetLeuAsnProLysMetCysValAlaGlyProAsnGlyProProGlyPheAsnPr 1088
QY 5328 TAATTCCCCCAATGGTGGATTAAAGAGAAATTCATAGGGTCTGGCTGCACAAA 5387
Db 1088 oAsnSerProAsnGlyGlyLeuArgGlnAsnSerIleGlySerGlyCysGlySerAlaAs 1108
QY 5388 CTCTTCAAACCTTTCAAGGGGTGTTCCACTGTGTCAGGAATGATGGGTGCAATGCCAGT 5447
Db 1108 nSerSerAsnPheGlnGlyValValProProGlyAlaArgMetMetGlyArgMetProVa 1128
QY 5448 CAATTTTGGTTCGAATTTCAATCCGAATATTCAAGTAAAGGCGAGTACCCCAACACCAT 5507
Db 1128 lAsnPheGlySerAsnPheAsnProAsnIleGlnValLysAlaSerThrProAsnThrI 1148
QY 5508 ACAATACATGCCAGTAAAGGCACAGAACGCCCAACAACATAACAATGAGCTAATAA 5567
Db 1148 eGlnTyrMetProValArgAlaGlnAsnAlaAsnAsnAsnAsnAsnGlyAlaAsnAs 1168
QY 5568 TGTGCGAATGCCACCTAGTCTGGAATTTTGCAGAGGTACGCTAACCTCAATGGGTGC 5627
Db 1168 nValArgMetProProSerLeuGlnPheLeuGlnArgTyrAlaAsnProGlnMetGlyAl 1188
QY 5628 TGTAGGCAATGGGTGCGCAATATGCCACCATCAGCCAGCGGACTCTGGAATGCC 5687
Db 1188 aValGlyAsnGlySerProIleCysProProSerAlaSerAspGlyThrProGlyMetPr 1208
QY 5688 AGGATTGATGCGGGGACCAAGAGCCGAGGTATGCTAATGAATTTCCGGAGAGCAACA 5747
Db 1208 oGlyLeuMetAlaGlyProGlyAlaGlyGlyMetLeuMetAsnSerSerGlyGlnH 1228
QY 5748 CCAGAACAAAGATCACAACAATCCTGGGCAAGCAATGATTAATCTTTCAGAATTG 5807
Db 1228 sGlnAsnLysIleThrAsnAsnProGlyAlaSerAsnGlyIleAsnPhePheGlnAsnC 1248
QY 5808 CAATCAATGTCTATTGTTGACGAAGAGGTGATTACC CGGCATGCGATCAATGAA 5867
Db 1248 sAsnGlnMetSerIleValAspGlnGlnGlyLeuProGlyHisAspGlySerMetAs 1268
QY 5868 TATTGTCAACCATCTATGATAAGGGGATGCGTCCACATGCCATGCCCAATGTAAT 5927
Db 1268 nIleGlyGlnProSerMetIleArgGlyMetArgProHisAlaMetArgProAsnValMe 1288
QY 5928 GGGTGGCGGATGCCACCGGTTAACAGGCAAAATTCAGTTTGACACAGTATCGGATGTTAT 5987
Db 1288 tGlyAlaArgMetProProValAsnArgGlnIleGlnPheAlaGlnSerSerAspGlyI 1308
QY 5988 TGACTGTGTGGGGATCCGTCATCATTTTTCATAAGCTTCTTGCAACAGCGCTGAGCC 6047
Db 1308 eAspCysValGlyAspProSerSerPhePheThrAsnAlaSerCysAsnSerAlaGlyPr 1328
QY 6048 ACACATGTTTGGATCAGACACAACAGGCCAATCAGCCTTAAGACACACAATPAAGAACAT 6107
Db 1328 oHisMetPheGlySerAlaGlnGlnAlaAsnGlnProLysThrGlnHisIleLysAsnI 1348
QY 6108 ACCTAGTGAATGTGTCAAAAACCAATCGGACTTGCAGTGGCAGACAGGCGAGATCCAAC 6167
Db 1348 eProSerGlyMetCysGlnAsnGlnSerGlyLeuAlaValAlaGlnGlnIleGlnLe 1368

QY 3408 GATTTAATGATGGGATGTCCGGTGGCGAGCGGATCTATTATTAATCCGACGATGCGA 3467
|||||
Db 561 AspleuemetGlymetSerGlyGlyGlySerIleIleasnProThrmelarG 580
QY 3468 CAACCTGATATGCCAGGTAAACGCCAAATCCGAGCTCTTATCGGCCGACAAGTTCAGACTT 3527
|||||
Db 581 GlnleuHleMetProGlyAsnAlalySerGlnleuLeuSerAlaThrSerSerGlyLeu 600
QY 3528 TCGGAAGATGTAATGCATCCAGGGGATGTTATATCAGATATGGTCCGTAATAGATGT 3587
|||||
Db 601 SerGluAspValMetHisProGlyAspValIleSerAspMetGlyAlaValIleGlyCys 620
QY 3588 AATAATATCAAAAAACCCAGTGTCAATGTGGATCTGGAGTAGGTGTTGTCACTGGAACA 3647
|||||
Db 621 AsnAsnAsnGlnLysThrSerValGlnCysGlySerGlyValGlyValIvalThrGlyThr 640
QY 3648 ACTGCAGCTGGAGTAAATGTCAATATGCAATTGCTCAAGCTCCGGCGCCCGAATGGCAAT 3707
|||||
Db 641 ThrAlaAlaGlyValAlaSnValAsnMetHisCysSerSerSerGlyAlaProAsnGlyAsn 660
QY 3708 ATGATGGGAAGCTCTACGATATGTAGCTTCGTTTGCAACAACAAGCTGCAACGTCAATC 3767
|||||
Db 661 MetMetGlySerSerThrAspMetLeuAlaSerPheGlyAsnThrSerCysAsnValIle 680
QY 3768 GGAACGGCCCCAGATATGTCTAAGGAAGTTTAAATCAAGATAGCCGAACCCATTACAT 3827
|||||
Db 681 GlyThrAlaProAspMetSerLysGlnValleuAsnGlnAspSerArgThrHisSerHis 700
QY 3828 CAAGGGGAGTTGCTCAATGGAGTGGTCCGAAGATTCAACATCAATTTTTCGAAGAACGC 3887
|||||
Db 701 GlnGlyGlyValAlaGlnMetGlnTyrSerLysIleGlnHisGlnPhePheGlnGluArg 720
QY 3888 CTCAGGGGGGCAAGCCCGACAGACAGTCACTGGAACTGTAGTACCACAACAGCAAAACCCCT 3947
|||||
Db 721 LeuLysGlyGlyLysProArgGlnValThrGlyThrValIvalProGlnGlnGlnThrPro 740
QY 3948 TCTGATCTGTGAAAACCTCGTTAAACAACAGGTGCGACCCCTGCAAGTCCACCTCCT 4007
|||||
Db 741 SerGlySerGlyGlyAsnSerLeuAsnAsnGlnValArgProLeuGlnGlyProProPro 760
QY 4008 CCTTACCACCTCATCCAGAGATCTGCGTCACTAACCATAGCCACTCAATCGCCCATCCC 4067
|||||
Db 761 ProTyrHisSerIleGlnArgSerAlaSerValProIleAlaThrGlnSerProAsnPro 780
QY 4068 TCGAGTCCAACAATCTATCTCTCCCGTCAACCCGCGACAACCGCAGCATGTGGATTG 4127
|||||
Db 781 SerSerProAsnAsnLeuSerLeuProSerProArgThrThrAlaAlaValMetGlyLeu 800
QY 4128 CCGACCAACTCTCTAGCATGGATGGAACAGATCATTTATCTGGATCTGTCGCAAGCT 4187
|||||
Db 801 ProThrAsnSerProSerMetAspGlyThrGlySerLeuSerGlySerValProGlnAla 820
QY 4188 AATACTTCGACGGTTCAGGCAGGCACAACAAGTGTCTCAGCAAAACAAGACTGTTT 4247
|||||
Db 821 AsnThrSerThrValGlnAlaGlyThrThrThrValleuSerAlaAsnLysAsnCysPhe 840
QY 4248 CAGGCAGACACCCCATCGCGTCAAAATCGTAGTAGAAATACCGGATCGTCAAGC 4307
|||||
Db 841 GlnAlaAspThrProSerProSerAsnGlnAsnArgSerArgAsnThrGlySerSerSer 860
QY 4308 GTTCTTACGCACTTAAGCAGCAACCCCAAGTACCCCCCTATCTCATCTATCCCCAAG 4367
|||||
Db 861 ValLeuThrHisAsnLeuSerSerAsnProSerThrProLeuSerHisLeuSerProLys 880
QY 4368 GAATTTGAGTCTTTCCGTCAGTCTCTGCTGTGTATGTTATTTGTTAAATTTTTTAA 4427
|||||
Db 881 GluPheGlnSerPheGlyGlnSerSer----- 889
QY 4428 GACAAATCAAAATATGAATTGCGTTAATATAGTTATATATTACATAACTCGGAAATTTG 4487
889 ----- 889

QY 4488 ATAGAAAAAATCAGGAATAGAAAAAATAATTATTTTCCGGACCGCCCATCTTCTTG 4547
889 ----- 889
QY 4548 AATCCAATTTCTGGAGTGATTTGTAGAGATATCTACTATTAAAAATTAACACGAATAAT 4607
889 ----- 889
QY 4608 CATATCCGTTAATTGAAATCACTATTGTTTAAATAGAAAATTAATAATATGTTATTATA 4667
889 ----- 889
QY 4668 ATATTCTTACAGGTATAACATGAAGTAGGCGACCAAGCCACAGGGTCAGCGGTAC 4727
890 -----Ala--GlyAspAsnMetLysSerArgArgProSerProGlnGlyGlnArgSerP 907
QY 4728 CAGTAAATAGTCTAATAGAGGCAATAAAGATGTACGATTGCTGCATCCAGTCCGTGTT 4787
|||||
Db 907 rovalAsnSerLeuIleGlnAlaAsnLysAspValArgPheAlaAlaSerSerProGlyP 927
QY 4788 TTAACCCGCATCCACATATGCAAAACCAATTCAATTTCAGCATTTAAACGCTATAAAATG 4847
|||||
Db 927 heasnProHisProHisMetGlnSerAsnSerAsnSerAlaLeuAsnAlaTyrLysMetG 947
QY 4848 GCTTACCAATATACAGATGGAGGTAAATATTAAATATTTTAAACGTTTGTGTT 4907
947 LysThrAsnIleGlnMetGlu--Arg----- 955
QY 4908 AATTTATCTTCTTTTTCAGCGCTCAAGCATCAGCGCAAGGTGGATCCGTACAATTTAGTCG 4967
956 -----GlnAlaSerAlaGlnGlyGlySerValGlnPheSerAr 968
QY 4968 GCGCTCCGATATATTTCCGCTAAATCCCAATAGTGGCAATCGGCCGCCACAACAAGAT 5027
|||||
Db 968 GArgSerAspAsnIleProLeuAsnProAsnSerGlyAsnArgProProProAsnLysMe 988
QY 5028 GACCCAACCTTCGATCCAATCTCTTTGGCACAACAATGCCCAACAATAACAAGTTG 5087
|||||
Db 988 tThrGlnAsnPheAspProIleSerSerLeuAlaGlnMetSerGlnGlnLeuThrSerCy 1008
QY 5088 CGTGTCCAGCATGGGTAGTCCAGCCGGAACCTGCTGTATGACATGATGGGGGTCCGGG 5147
|||||
Db 1008 sValSerSerMetGlySerProAlaGlyThrGlyGlyMetThrMetMetGlyGlyProG 1028
QY 5148 ACCGTCCGACATCAATATTGAGCATGGAATATTTCCGGACTAGATGATCAGGAATAGA 5207
|||||
Db 1028 YProSerAspIleAsnIleGlnHisGlyIleIleSerGlyLeuAspGlySerGlyIleAs 1048
QY 5208 TACCATAAATCAAAATACTGTTCATTCATGAATGTCGTAATGAACCTCAATGGGTCCCG 5267
|||||
Db 1048 pThrIleAsnGlnAsnAsnCysHisSerMetAsnValIalMetAsnSerMetGlyProAr 1068
QY 5268 AATGCTGAATCCTAAATGTGCGTAGCAGCGGCTCCAATGAGCCGCTGCTTAATCC 5327
|||||
Db 1068 gMetLeuAsnProLysMetCysValAlaGlyGlyProAsnGlyProProGlyPheAsnPr 1088
QY 5328 TAATTCCTCCCAATGTGTGATTAAAGAGAATTCATAGGGTCTGGCTGTGCTCAGCAA 5387
|||||
Db 1088 oAsnSerProAsnGlyGlyLeuArgGluAsnSerIleGlySerGlyCysGlySerAlaAs 1108
QY 5388 CTCTTCAAACTTTCAAGGGGTGTTCCACCTGCTGTCAGAAATGATGGGTGGAATGCCAGT 5447
1108 nSerSerAsnPheGlnGlyValIvalProProGlyAlaArgMetMetGlyArgMetProva 1128
QY 5448 CAATTTGCTTGAATTTCAATCCGAATATTCAAGTTAAAGCGAGTACCCCAACACCAT 5507
1128 lAsnPheGlySerAsnPheAsnProAsnIleGlnValLysAlaSerThrProAsnThrIl 1148
QY 5508 ACAATPACATGCCAGTAAGGGCAGAGAACGCCAACAACAATAACAATGAGACTAATAA 5567
|||||
Db 1148 eGlnTyrMetProValArgAlaGlnAsnAlaAsnAsnAsnAsnAsnGlyAlaAsnAs 1168
QY 5568 TGTGCGAATGCCACCTAGTCTGGAATTTTTCAGAGAGGTACGCTAACCTCAATGGGTGC 5627

Db 1168 nValArgMetProProSerLeuGluPheLeuGlnArgTyrAlaAsnProGlnMetGlyAl 1188
QY 5628 TGTAGGCAATGGGTCCCAATATATGCCCCACCATCAAGCGACGGTACTCTCGAATGCC 5687
Db 1188 aValGlyAsnGlySerProIleCysProProSerAlaSerAspGlyThrProGlyMetPr 1208
QY 5688 AGGATTGATGGCGGGAACGAGAGCCGGAGGTATGCTAATGAATTCCTCCGGAGACAACA 5747
Db 1208 cGlyLeuMetAlaGlyProGlyAlaGlyGlyMetLeuMetAsnSerSerGlyGluGlnHi 1228
QY 5748 CCAGAACAGATCACAAACAATCCTGGGGCAAGCATGTATTAACTTCTTCAAGAATTG 5807
Db 1228 sGlnAsnLysIleThrAsnAsnProGlyAlaSerAsnGlyIleAsnPheGlnAsnCyl 1248
QY 5808 CAATCAATGTCTATTGTGACGAGAGGGGTGATTACCCGGCCATGACGATCAATGAA 5867
Db 1248 sAsnGlnMetSerIleValAspGluGluGlyLeuProGlyHisAspGlySerMetAs 1268
QY 5868 TATTGGTCAACCATCTATGATAAGGGCATGCCCTCCACATGCCATGGCGCCAAATGTAAT 5927
Db 1268 nIleGlyGlnProSerMetIleArgGlyMetArgProHisAlaMetArgProAsnValMe 1288
QY 5928 GGGTGGCGGATGCCACCCGTTAACAGAGCAAAATTCAGTTTGACACAGTCATCGATGTTAT 5987
Db 1288 tGlyAlaArgMetProProValAsnArgGlnIleGlnPheAlaGlnSerSerAspGlyI1 1308
QY 5988 TGACTGTGTGGGGATCCGTCATCATTTTTCATTAACGCTTCGCAACAGCGCTGGACC 6047
Db 1308 eAspCysValGlyAspProSerSerPhePheThrAsnAlaSerCysAsnSerAlaGlyPr 1328
QY 6048 ACACATGTTTGATCAGACACAACAGGCCAATCAGCCTAAGACACAACACATTAAGAATCAT 6107
Db 1328 cHisMetPheGlySerAlaGlnGlnAlaAsnGlnProLysThrGlnHisIleLysAsnI1 1348
QY 6108 ACCTAGTGAATGTGTCAAAAACCAATCGGGACTTGCAAGTGCGACACAGGGCAGATCCAAT 6167
Db 1348 eProSerGlyMetCysGlnAsnGlnSerGlyLeuAlaValAlaGlnGlyGlnIleGlnLe 1368
QY 6168 GCATGGGCAAGGACATGCGCAGGGTCAGTCTTTAATTGGACCTACTAATTAATTTAAT 6227
Db 1368 uHisGlyGlnGlyHisAlaGlnGlnGlnSerLeuIleGlyProThrAsnAsnLeuMe 1388
QY 6228 GTCAACTGCCGAAGTGTCAAGTGTCTACTAAACGCTGCTCGCATCAATTTGTAGGTCC 6287
Db 1388 tSerThrAlaGlySerValSerAlaThrAsnGlyValSerGlyIleAsnPheValGlyPr 1408
QY 6288 CTCTTCTACGACCTGAAGTATGCTCCAGCAATATCATAGTTTTCAGCAGAGTTATATGC 6347
Db 1408 cSerSerThrAspLeuLysTyrAlaGlnGlnGlnIleGlnHisSerPheGlnGlnLeuTyrAl 1428
QY 6348 TACCAACACCAAGATCAACAACAACAGCATATGCACAGCAGCAGCAGCAACATGAT 6407
Db 1428 aThrAsnThrArgSerGlnGlnGlnGlnHisMetHisGlnGlnHisGlnSerAsnMetI1 1448
QY 6408 AACAAATGCCCGCAATTTATCACCAAAATCCAACGTTCTTTGTCAACAACA 6456
Db 1448 eThrMetProProAsnLeuSerProAsnProThrPhePheValAsnLys 1464
RESULT 4
ABB58779
ID ABB58779 standard; protein; 1429 AA.
XX ABB58779;
AC ABB58779;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 3129.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
OS Drosophila melanogaster.

XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
PI
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL02882.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX
PS Disclosure; SEQ ID NO 3129; 21bp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1429 AA;
SQ
Alignment Scores:
Pred. No.: 0 Length: 1429
Score: 7040.00 Matches: 1425
Percent Similarity: 71.09% Conservative: 1
Best Local Similarity: 71.04% Mismatches: 3
Query Match: 58.49% Indels: 577
DB: 4 Gaps: 6
US-10-664-859-1 (1-6909) x ABB58779 (1-1429)
QY 453 ATGCTCTCGACACAATGCCCGGACGTCCAAACCAACAGCCGCAACCAACTCCGAT 512
Db 1 MetLeuSerThrThrMetProArgSerProThrGlnGlnGlnProGlnProAsnSerAsp 20
QY 513 GCCTCCTCAACAAGTGCATCTGATCAAACTCTGGAGCAGCGATCGAAATGGGGACTCG 572
Db 21 AlaSerSerThrSerAlaSerGlySerAsnProGlyAlaAlaIleGlyAsnGlyAspSer 40
QY 573 GCGGCGAGCAGAAGTTCTCCGAAGACCCTTAATAGCGAACCCTTTCTACTTGTGCGCG 632
Db 41 AlaAlaSerArgSerSerProLysThrLeuAsnSerGluProPheSerThrLeu-Ser-- 59
QY 633 GGTAGACTTGATTTGATTCTCTTTGTCGGAATTATACAACCTTCTGTGTTCCAGA 692
Db 60 -----ProAs 61
QY 693 TCAATTAATTTGACGCCGAGAAGAGCGACTGAGAAAGCGGACTATCACTAGTGATAA 752
Db 61 pGlnIleLysLeuThrProGluGluGlyThrGluLysSerGlyLeuSerThrSerAspLys 81
QY 753 AGCTGCCACTGAGAGAGCCCGCAGGCAAGTGAATAATATCTGCCCGAGGACAAACTATGCT 812
Db 81 sAlaAlaThrGlyGlyAlaProGlySerGlyAsnAsnLeuProGluGlyGlnThrMetLe 101
QY 813 AAGGCAGAACTCTACGAGCACAATCAACTCGTGCCTAGTGGCTTCTCCACAAAACCTCCAG 872
Db 101 uArgGlnAsnSerThrSerThrIleAsnSerCysLeuValAlaSerProGlnAsnSerSe 121

QY 873 TGAACACTGCAATAGCAGCAATGTGTCTCTACAGTGGCCTTACTCAGATGTAGATTG 932
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
121 rGlHISerASnSerSerASnValSerAlaThrValGlyLeuThrGlnMetValAspCy 141
QY 933 TGACGAGCAATCGAAGAAAAACAATGTAGTGTGAAGGACGAGGAACTGGTAAAGACTGC 992
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
141 bAspGluGlnSerLysLysASnLysCySerValLysAspGluGluAla----- 157
QY 993 CCTACAAATGGTTTAAATTTTAAATGTATGGCGGTTACCGTTTGTTAATCATTTAATT 1052
Db 157 ----- 157
QY 1053 GTTTTGTCTACTACTACAAATTTAGTTTAAACTTGTAACCTGACTAAAACCTCG 1112
Db 157 ----- 157
QY 1113 CGAAGCTCGATCAAAACAGACATTTTCTGGAAACCGTAATTAAAGCTCATAAAATATTAA 1172
Db 157 ----- 157
QY 1173 ATTCACTTGAATGCAATGCAATATCATATGTACTCAAAACATCTCAAGAAAGACCTCAA 1232
Db 157 ----- 157
QY 1233 TTGGATCACTAATTAGTTTGAGAAAAAATTGCTGTACTTTTAAAGATATATTATTAA 1292
Db 157 ----- 157
QY 1293 AAATTTGCTGAGTAATGATATATATAGTCACAATAATTTTGTAACTGCTAAAGCA 1352
Db 157 ----- 157
QY 1353 TTTTGAATAGCCGTGCTACGAGATGCTACTAGACGCGGTGAAGCTAATTTTATT 1412
Db 157 ----- 157
QY 1413 AAAAGCTGCTAATATTCATTAACCATTAATGTCCCATTTAGAAATAAGTTCTAATA 1472
Db 158 -----GluIleSerSerASnLys 163
QY 1473 AGCAAAAGGTCAAGCAGCTGTGCGGCTGCGAAACAGTTTCTACATCCAGTTTGACTGT 1532
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
163 sAlaLysGlyGlnAlaAlaGlyGlyCySerGluThrGlySerThrSerSerLeuThrVa 183
QY 1533 CAAGAAGAACCCACCGATGTCTTAGGCAAGTTTAGTAATATGAAAAAAGAAAGAGA 1592
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
183 lLysGluGlnProThrAspValLeuGlySerLeuValAsnMetLysGlyGluArgGly 203
QY 1593 AAATCATTCGCCAACGATGTCCCTGTGGTTTGGTTCAATTGGTAAATGCACAGACAA 1652
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
203 uASnHISerProThrMetSerProValGlyPheGlySerIleGlyASnAlaGlnAspAs 223
QY 1653 CTCCGCTACACCGGGTAAGTTTAAAGATCCATATTAAGCAAAATAACAAGAATTATGT 1712
Db ||||||||||nSerAlaThr----- 226
QY 1713 CAGTTACCAATTTATTGTAGTCAAAAGACTACTATAGCGATATCTCTGCTTTTAA 1772
Db 226 ----- 226
QY 1773 TTTTATTTTAATAGAAATACGAATATTCTAATTGTAAATAAATTGATTAAATTAA 1832
Db 226 ----- 226
QY 1833 CTAGAATTTAAAACTTTTGAATTAGACATACCCTTCCAAAAATCAGTAATCATTGGG 1892
Db 226 ----- 226
QY 1893 AACGAGAGTGTGTCCCGAAGAGACTACTATAAAACCTTTGAGCTATCTGATCTGCA 1952
Db 226 ----- 226

QY 1953 CGCTACTAAAAATGATTAGTTTAGAAAAATGGGTGTAAATTTTGTAGGAAGTTTTCATTTT 2012
Db 226 ----- 226
QY 2013 AGAAGAAATGTGATTATTTATTAACCCCTTCAAGCGGAATACTATTGTTCTACGATA 2072
Db 226 ----- 226
QY 2073 TTTTGAAAAAACAATGGTTAAGTTGGAAGTGCCCTATTAACAGAAATTCACCGGTTTCA 2132
Db 226 ----- 226
QY 2133 AATACTAACCGGTTTGTGATTTAATTTTGTATTAAATGAGAAATTATCACACCTTCAGTTA 2192
Db 226 ----- 226
QY 2193 AAATGTTTAATTGATTAAAGTCGACAATCACAGACAGATTTCATTTTTCGTGTATAT 2252
Db 226 ----- 226
QY 2253 ATAGAAGTCGCCCTTACACTCTTCTGGCGCGCTTACCACCTACGTGAGTCCGCCCGCA 2312
Db 226 ----- 226
QY 2313 GTGATTTATATAGATGATTAGCAGTTATTAAATTTTATAGGTATTTTAATAATATAT 2372
Db 226 ----- 226
QY 2373 CTATATTATTCATTTTACATAGTTTAAATTTGAAGAATTTCAAACGACAGTACCAGGAA 2432
Db 227 -----Pro--ValLysIleGluArgIleSerASnAspSerThrThrGlu 240
QY 2433 AAAAAAGATCGTCCCTTGACAAATGAATATGACGAAATGAGATGGAAGGCTGCAATCAG 2492
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
241 LysLysGlySerSerLeuThrMetASnASnAspGluMetSerMetGluGlyCyASnGln 260
QY 2493 TTGAATCCCGATTTTATCAATGAATCTTTAAATAATCCGTGCAATTCGAGCATATTAGTA 2552
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
261 LeuAsnProAspPheIleAsnGlnSerLeuASnAsnProAlaIleSerSerIleLeuVal 280
QY 2553 AGCGGATGAGACCAATACCCGGAATCGGAGTTGGAGCGGGGACGGAAATTATTGACT 2612
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
281 SerGlyValGlyProIleProGlyIleGlyValGlyAlaGlyThrGlyASnLeuLeuThr 300
QY 2613 GCCAAGCCCAATGGAATCTCCTGGGTAAGCAGTAATGTTGGAATTACATGCACACAGCAA 2672
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
301 AlaAsnAlaAsnGlyIleSerSerGlySerSerASnCyLysLeuAspTyrMetGlnGlnGln 320
QY 2673 AATCACAATATTCGTGTTTCAACTCAGCTGCGCAACAAGGGCCGAATCAGTTTAAAGC 2732
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
321 ASnHISilePheValPheSerThrClnLeuAlaAsnLysGlyAlaGlnSerValLeuSer 340
QY 2733 GGTCAATTTCAAACTATATTGCGGTATCAGCTGCAGCTCAGCTCTTACAAAAAGCTTCTGTG 2792
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
341 GlyGlnPheGlnThrIleIleAlaTyrHISCySerThrGlnProAlaThrLysSerPheLeu 360
QY 2793 GAAGACTTTTATGAAAAACCTTTTAAAGATTAAACAAGTTACAGCGGCACAAATTCGGTC 2852
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
361 GluAspPhePheMetLysASnProLeuLysIleAsnLysLeuGlnArgHISAsnSerVal 380
QY 2853 GGTATGCCATGGATTAGGCATGGGGCAGGTTGGACTACTCTCTTAATCTGTAGCCAAA 2912
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
381 GlyMetProTyrIleGlyMetGlyGlnValGlyLeuThrProProAsnProValAlaLys 400
QY 2913 ATAACAACAAGCAGCCACATACAAAGACCGTAGGCTTATGAAACCCCAATTCAATCAA 2972
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
401 IleThrGlnGlnGlnProHISThrLysThrValGlyLeuLeuLysProGlnPheAsnGln 420
QY 2973 CATGAAAACAGCAACGTAAGTACTGTAAAGCGCGCTAGCAACTCTTTGTGCAACAGTCT 3032
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
421 HISGluAsnSerLysArgSerThrValSerAlaProSerASnSerPheValAspGlnSer 440
QY 3033 GATCCTATGGGCAACGAACGTGAATTGATGTGCTGGAAAGCGGATTCCTCAAAACACAGT 3092

Db 441 AspPrometGlyAsnGluThrGluLeuMetCysTrpGluGlySerSerAsnThrSer 460
QY 3093 AGGTCTGGACAAAACTCAGGAATCATGTAGACAGTATCAGTACATCCAGCGAGTCACAG 3152
Db 461 ArgSerGlyGlnAsnSerArgAsnHisValAspSerIleSerThrSerSerGluSerGln 480
QY 3153 GCAATAAAGATACTGGAAGCAGCTGGCGTTGATTGGGACAGGTCACAAAAGAAAGCGAT 3212
Db 481 AlaIleLysIleLeuGluAlaIaGlyValAspLeuGlyGlnValThrLysGlySerAsp 500
QY 3213 CCTGGCCCTGACAACTGAAAAACAACATTTGATCTGCAAGAGGTTAAGGTTCCAGACGAA 3272
Db 501 ProGlyLeuThrThrGluAsnAsnIleValSerLeuGlnGlyValLysValProAspGlu 520
QY 3273 AACCTTACACCAACAGCGGCAACATCGGAGAGAACAGTTGGCAAAAATAAAAATG 3332
Db 521 AsnLeuThrProGlnGlnArgGlnHisArgGluGluGlnLeuAlaLysIleLysLysMet 540
QY 3333 AATCAATTTCTTTTCTCTGAAAAATGAGAATTCAGTAGGAGCTAATGTAAGCTCACAGATA 3392
Db 541 AsnGlnPheLeuPheProGluAsnGluAsnSerValGlyAlaAsnValSerSerGlnIle 560
QY 3393 ACAAATAATTCAGAGATTTAATGATGGGATGTCGGGTGGCGGAGCGGATCTATTATA 3452
Db 561 ThrLysIleProGlyAspLeuMetMetGlyMetSerGlyGlyGlySerIleIle 580
QY 3453 AATCCGACGATGCACAACCTGCATATGCCAGGTAAAGCCAAATCGAGCTCTTATCGGCG 3512
Db 581 AsnProThrMetArgGlnLeuHisMetProGlyAsnAlaLysSerGluLeuLeuSerAla 600
QY 3513 ACAAGTTCAGACTTTCGGAAGATGTAATGCATCCAGGGATGTTATATCAGATATGGGT 3572
Db 601 ThrSerSerGlyLeuSerGluAspValMetHisProGlyAspValIleSerAspMetGly 620
QY 3573 GCCGTATAGGATGTAATAATAATCAAAAAACAGTGTGCAATGTGATCTGAGTAGT 3632
Db 621 AlaValIleGlyCysAsnAsnAsnGlnLysThrSerValGlnCysGlySerGlyValGly 640
QY 3633 GTTGTCACTGGAACAACTGCAGCTGGAGTAAATGTCATATGCAATTGCTCAAGCTCCGGC 3692
Db 641 ValValThrGlyThrThrAlaIaGlyValAsnValAlaSerMetHisCysSerSerSerGly 660
QY 3693 GCCCCGAATGGCAATATGATGGGAAGCTCTACGATATGCTAGCCCTGTTGGCAACACA 3752
Db 661 AlaProAsnGlyAsnMetMetGlySerSerThrAspMetLeuAlaSerPheGlyAsnThr 680
QY 3753 AGTCGAACGTCAATCGGAACGGCCCAAGATATGCTTAAGGAAGTTTAAATCAAGATAGC 3812
Db 681 SerCysAsnValIleGlyThrAlaProAspMetSerLysGluValLeuAsnGlnAspSer 700
QY 3813 CGAACCCATTCACATCAAGGGGGAGTTGCTCAATGAGTGTGGAAGATTCACATCAA 3872
Db 701 ArgThrHisSerHisGlnGlyValAlaGlnMetGluTrpSerLysIleGlnHisGln 720
QY 3873 TTTTTCGAAGAACCGCTCAAGGGGGCAAGCCCAACAAGTCACTGGAATCTGATACCA 3932
Db 721 PhePheGluGluArgLeuLysGlyGlyLysProArgGlnValThrGlyThrValValPro 740
QY 3933 CAACAGCAAAACCCCTTCTGGATCTGGTGAAACTGTTAAACAACAGGTGCGACCCCTG 3992
Db 741 GlnGlnGlnThrProSerGlySerGlyLysAsnSerLeuAsnAsnGlnValArgProLeu 760
QY 3993 CAAGGTCCACCTCTCTTACCACCTCCATCCAGAGATCTGCTCAGTACCAATAGCCACT 4052
Db 761 GlnGlyProProProGlyThrHisSerIleGlnArgSerAlaSerValProIleAlaThr 780
QY 4053 CAATGCCCAATCCCTCGAGTCCAAACAATCTACTCTCCGCTCACCGCGGACAACCGCA 4112
Db 781 GlnSerProAsnProSerSerProAsnAsnLeuSerLeuProSerProArgThrThrAla 800
QY 4113 GCAGTCATGGGATTCGCGCACTCTCTAGCATGGATGGAACAGATCATATCTTGA 4172

Db 801 AlaValMetGlyLeuProThrAsnSerProSerMetAspGlyThrGlySerLeuSerGly 820
QY 4173 TCTGTTCCGCAAGCTAATACTTCGACGGTTACAGGACGACACAAACAGTGTCTCAGCA 4232
Db 821 SerValProGlnAlaAsnThrSerThrValGlnIaGlyThrThrThrValLeuSerAla 840
QY 4233 AACAGAAGCTTTTTCAGGCAGACACCCCATCGCCGTCAAATCAAATCGTAGTAGAAAT 4292
Db 841 AsnLysAsnCysPheGlnAlaAspThrProSerProSerAsnGlnAsnArgSerArgAsn 860
QY 4293 ACCGATCGTCAAGCGTTCTTACGATTAATTAAGCAGCAACCCCAAGTACCCCTTATCT 4352
Db 861 ThrGlySerSerSerValLeuThrHisAsnLeuSerSerAsnProSerThrProLeuSer 880
QY 4353 CATCTATCCCCAAAGGAATTGAGTCTTTCGGTCAGTCCCTCTGCTGTTATTTG 4412
Db 881 HisLeuSerProLysGluPheGluSerPheGlyGlnSerSer----- 894
QY 4413 TTTAATTTTAAAGACAATCAATATGAAATTGCGTTAATTAATTAATATTTACA 4472
Db 894 ----- 894
QY 4473 TAACTCGGAATTTGATAGAAAAAATCAGGAATAGAAAAAATAATTATTTCCGGAACG 4532
Db 894 ----- 894
QY 4533 CCCATCATTTCTGAATCCAATTTCTGAGTGATTTGTAAGATTAATCTATTTAAAA 4592
Db 894 ----- 894
QY 4593 TTAACACGAAAAATTCATATCCGTTAATGAAATCACTATTGTTAATGAATAATTAA 4652
Db 894 ----- 894
QY 4653 AATATGTTTATTAATAATTTCTACAGTGATTAACATGAAAAGTAGCGCAACCAAGCCAC 4712
Db 895 -----Ala--GlyAspAsnMetLysSerArgArgProSerProG 907
QY 4713 AGGTCAGCGGTCAACAGTAATAGTCTAATAGAGCAAAATAAGATGTACGATTTGCTG 4772
Db 907 LngLysGlnArgSerProValAsnSerLeuIleGluAlaAsnLysAspValArgPheAla 927
QY 4773 CATCCAGTCCGTGTTTAAACCCGCATCCACATATGCAAAAGCAATTCAACATTTAA 4832
Db 927 IaSerSerProGlyPheAsnProHisProHisMetGlnSerAsnSerAsnSerAlaLeu 947
QY 4833 ACGCTATTAATGGGCTCTACCAATATACAGATGGAGGTAAATATTTAATTTTATT 4892
Db 947 snAlaTyrlLysMetGlySerThrAsnIleGlnMetGlu--Arg----- 960
QY 4893 TTAACGTTTTGTGTTAATTATCTTTTTCAGCGTCAAGCATCAGCGCAAGGTGATC 4952
Db 961 -----GlnAlaSerAlaGlnGlyLys 968
QY 4953 CGTACAATTTAGTCGGCGCTCCGATAATATTCGCTAAATCCCAATAGTGCATCGGCC 5012
Db 968 rValGlnPheSerArgArgSerAspAsnIleProLeuAsnProAsnSerGlyAsnArgPr 988
QY 5013 GCCACCAACAGATGACCCCAAACTTCGATCCAAATCTCTTCTTGGCAAAATGTCCCA 5072
Db 988 OProProAsnLysMetThrGlnAsnPheAspProIleSerSerLeuAlaGlnMetSerG1 1008
QY 5073 ACACTAACAAGTTGCGTGTCCAGCATGGGTAGTCCAGCCGGAACGTGGTATGACGAT 5132
Db 1008 nglnLeuThrSerCysValSerSerMetGlySerProAlaGlyThrGlyGlyMetThrMe 1028
QY 5133 GATGGGGGTCCGGGACCGTCCGACATCAATATTGAGCATGGAATATTTCGGGACTAGA 5192
Db 1028 tMetGlyGlyProGlyProSerAspIleAsnIleGluHisGlyIleIleSerGlyLeuAs 1048
QY 5193 TGATCAGGAATAGATACCATAATCAAAATAACTGTCAATTCAATGAATGTCGTAATGAA 5252
Db 1048 pGlySerGlyIleAspThrIleAsnGlnAsnAsnCysHisSerMetAsnValValMetAs 1068

OY 5253 CTCATGGGTCCCCGAATGCTGAATCCTAAATGTGCGTAGACGGCGTCCAAATGACC 5312
Db 1068 nSerMetGlyProArgMetLeuAnProLysMetCysValAlaGlyGlyProAnGlyPr 1088
OY 5313 GCCTGGCTTTAATCCTAATTTCCCCCAATGGTGATTAGAGAGAATTCATAGGGTCTGG 5372
Db 1088 oProGlyPheAnProAnSerProAnGlyGlyLeuArgGlyAnSerIleGlySerGI 1108
OY 5373 CTGTGGCTCAGCAAACTCTTCAAACTTTCAGAGGGTGTGTCCACTGTGTCCAGAAATGAT 5432
Db 1108 yCyseGlySerAlaAnSerSerAlaAnPheGlnGlyValValProGlyAlaArgMetMe 1128
OY 5433 GGGTCGAATGCCAGTCAATTTGGTTCGAATTTCAATCCGAATATTCAGGTAAAGCGCAG 5492
Db 1128 tGlyArgMetProValAnPheGlySerAnPheAnProAnIle----- 1143
OY 5493 TACCCCAACACCATATACATATGCGAGTAAAGGCACAGAACGCCAACAAATTAACAA 5552
Db 1143 ----- 1143
OY 5553 CAATGAGCTAATAATGTGCGAATGCCACTAGTCTGGAATTTTTCAGAGGTACGCTAA 5612
Db 1144 -----GlnArgTyrAlaAs 1148
OY 5613 CCCTCAATGGGTGCTGTAGGCAATGGGTGCCAATATATGCCCAACCATCAGCCAGCGACGG 5672
Db 1148 nProGlnMetGlyAlaValAlaGlyAnGlySerProIleCysProProSerAlaSerAspGI 1168
OY 5673 TACTCCTGGAATGCCAGGATTGATGGCGGACCAAGAGCCGGAGGTATGCTAATGAATTC 5732
Db 1168 yThrProGlyMetProGlyLeuMetAlaGlyProGlyAlaGlyGlyMetLeuMetAnSe 1188
OY 5733 TTCCGGAGAGCAACACCAAGACAAGATCACAAACAATCCTGGGGCAAGCAATGTATTAA 5792
Db 1188 rSerGlyGlnGlnHisGlnAnLysIleThrAnAnProGlyAlaSerAnGlyIleAs 1208
OY 5793 CTTCTTTCAGAATTGCAATCAAAATGCTATGTGTGACGAAGAGGGTGATTACCCGGCCA 5852
Db 1208 nPhePheGlnAnSncysAnGlnMetSerIleValAspGlnGlnGlyLeuProGlyH1 1228
OY 5853 TGACGGATCAATGAATATTGGTCAACCACTATGATTAAGGGGCAATGCTCCACATGCCAT 5912
Db 1228 sAspGIySerMetAnIleGlyGlnProSerMetIleArgGlyMetArgProHisAlaMe 1248
OY 5913 GCGGCCAATGTAATGGGTGCGCGGATGCCACCCGTTAACAGGCAAAATTCAGTTTGACA 5972
Db 1248 tArgProAnValMetGlyAlaArgMetProProValAnArgGlnIleGlnPheAlaGI 1268
OY 5973 GTCATCGGATGTATTGACTGTGTCCGGGATCCGTCATCATTTTTCACCTAACGCTTCCTG 6032
Db 1268 nSerSerAspGIyIleAspCysValGIyAspProSerSerPhePheThrAnAlaSerCy 1288
OY 6033 CAACAGCGCTGGACCACACATGTTGGATCAGCACACAGGCGCAATCAGCCTAAGACACA 6092
Db 1288 sAnSerAlaGIyProHisMetPheGlySerAlaGlnGlnAlaAnGlnProLysThrGI 1308
OY 6093 ACACATAAAGACATACCTAGTAGAATGTGTCAAAACCAATCGGACTTGACAGTGACACA 6152
Db 1308 nHisIleLysAnIleProSerGIyMetCysGlnAnSngInSerGIyLeuAlaValAlaGI 1328
OY 6153 AGGCGAGATCCAAGTGCATGGGCAAGGACATGCGCAGGTCAGTCTTTAATTGGACCTAC 6212
Db 1328 nGlyGlnIleGlnLeuHisGIyGlnGlyHisAlaGlnGlyGlnSerLeuIleGIyProth 1348
OY 6213 TAATAATAATTAAATGTCAACTGCCGGAAGTGTCAAGTCTACTTAACGGTGTCTTGGCAT 6272
Db 1348 rAnAnAnAnLeuMetSerThrAlaGlySerValSerAlaThrAnGlyValSerGIyI1 1368
OY 6273 CAATTTGTAAGTCCCTCTTCTACGGAGCTGAAGTATGCCAGCAATATCATAGTTTTC 6332
Db 1368 eAnnPheValGIyProSerSerThrAspLeuLysTyrAlaGlnGlnTyrHisSerPheGI 1388

OY 6333 GCAGCAGTTATATGCTTACCAACACCAAGTCAACACACAGCATATGCACCAGCAGCA 6392
Db 1388 nGlnGlnLeuTyrAlaThrAnThrArgSerGlnGlnGlnHisMetHisGlnGlnH1 1408
OY 6393 CCAGAGCAACATGATTAACAATGCCCGCGAATTATACCAAAATCCAAAGTCTTTGTCAA 6452
Db 1408 sGlnSerAnMetIleThrMetProProAnLeuSerProAnProThrPhePheValAs 1428
OY 6453 CAAA 6456
Db 1428 nLys 1429

RESULT 5

ABW01532
ID ABW01532 standard; peptide; 112 AA.
XX
AC ABW01532;
XX
DT 15-JAN-2004 (first entry)
XX
DE Drosophila species legless (lgs) peptide #5.
XX
KW legless protein; lgs; cell fate disorder; blood disease; gene therapy;
KW cancer; tissue regeneration; tissue repair; cytostatic.

XX Drosophila sp.
XX US2003114413-A1.
XX
PD 19-JUN-2003.
XX
PF 19-DEC-2002; 2002US-00322579.
XX
PR 28-JUL-2000; 2000US-0221502P.
PR 27-JUL-2001; 2001US-00915543.
XX
PA (UYZU-) UNIV ZURICH.

XX
PI Basler K, Brunner E, Froesch B, Kramps T, Peter O;
XX
XX
DR WPI; 2003-829432/77.
XX
PT Novel lgs polypeptide useful for isolation of lgs-binding proteins,
PT diagnosing disorders of cell fate, treating diseases such as cancer.
XX
PS Claim 28; Fig 7B; opp; English.
XX

XX The invention relates to novel legless (lgs) proteins and polynucleotides
CC encoding such proteins. Lgs sequences are useful for the treatment of
CC disorders of cell fate such as differentiation or proliferation. The
CC invention is used to treat blood disease or a cancerous condition
CC characterised by over-stimulation of the Wnt pathway such as colon,
CC breast, head and neck, brain, thyroid, medulloblastoma or skin cancer and
CC is administered to prevent progression from a pre-neoplastic or non-
CC malignant condition to a neoplastic or malignant state. It is
CC administered to promote tissue regeneration and repair. The invention is
CC also useful in the therapy of diseases cost by an over-activation of Wg
CC pathway. It is useful for reducing lgs gene expression in an invertebrate
CC or vertebrate organism or an invertebrate or vertebrate cell line. The
CC invention is also useful in gene therapy. The present sequence is
CC Drosophila species legless (lgs) peptide

XX
SQ Sequence 112 AA;

Alignment Scores:

Pred. No.: 9.01e-40 Length: 112
Score: 570.00 Matches: 112
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.74% Indels: 0
DB: 7 Gaps: 0

US-10-664-859-1 (1-6909) x ABW01532 (1-112)

QY 4029 TCTGCGTACGTACCAATAGCCACTCAATCGCCCAATCCCTCGAGTCCAAACAATCTATCT 4088
Db 1 SerAlaSerValProIleAlaThrGlnSerProAsnProSerSerProAsnAsnLeuSer 20
QY 4089 CTCCCGTACCCGGGACAAACCGGACGAGTCATGGGATTGCCGACCAACTCTCTTAGCATG 4148
Db 21 LeuProSerProArgThrThrAlaAlaValMetGlyLeuProThraAsnSerProSerMet 40
QY 4149 GATGAACAGGATCATTTATCTGGATCTGTTCCGCAAGCTAATACTTCGACGGTTGAGGCA 4208
Db 41 AspGlyThrGlySerLeuSerGlySerValProGlnAlaAsnThrSerThrValGlnAla 60
QY 4209 GGCACAACAACAGTGTCTTCAGCAACAAGAACTGTTTTCAGGACAGACACCCCATCGCCG 4268
Db 61 GlyThrThrThrValLeuSerAlaAsnLysAsnCysPheGlnAlaAspThrProSerPro 80
QY 4269 TCAATCAAAATCGTAGTAGAATATACCGGATCGTCAGCGTTCTTACGCATACTTAAGC 4328
Db 81 SerAsnGlnAsnArgSerArgAsnThrGlySerSerSerValLeuThrHisAsnLeuSer 100
QY 4329 AGCAACCCCAAGTACCCCTTATCTCATCTATCCCCA 4364
Db 101 SerAsnProSerThrProLeuSerHisLeuSerPro 112
RESULT 6
ADJ71898
ID ADJ71898 standard; protein; 112 AA.
XX
AC ADJ71898;
XX
DT 20-MAY-2004 (first entry)
XX
DE Fruit fly legless (lgs) peptide fragment #5.
XX
KM Fruit fly; legless; lgs; cell differentiation disorder;
KM cell proliferation disorder; cancer; wnt pathway; medulloblastoma; colon;
KM breast; head; neck; brain; thyroid; skin; blood disease;
KM tissue regeneration; tissue repair; cytoslatic.
XX
OS Drosophila melanogaster.
XX
PN US2004038901-A1.
XX
PD 26-FEB-2004.
XX
PF 22-SEP-2003; 2003US-00664859.
XX
PR 28-JUL-2000; 2000US-0221502P.
PR 27-JUL-2001; 2001US-00915543.
XX
PA (UYZU-) UNIV ZURICH.
XX
PI Basler K, Brunner E, Froesch B, Kramps T, Peter O;
XX
DR WPI; 2004-203288/19.
XX
PT Novel polypeptide sharing one or more homologue amino acid domains with
PT Legless protein being functional homologue of legless, useful for
PT diagnosing disorders of cell fate.
XX
PS Disclosure; SEQ ID NO 10; 62pp; English.
XX
CC The invention relates to a polypeptide sharing one or more homologous
CC amino acid domains with a legless (lgs) protein and is therefore a
CC functional homologue of lgs. The invention also relates to a nucleotide
CC sequence encoding a protein present in invertebrate and/or vertebrate
CC organisms, the nucleotide sequence encoding a protein comprising a
CC positive function in a regulatory pathway and the use of the polypeptide
CC for the isolation of lgs-binding proteins by carrying out an assay chosen
CC from an in vitro binding assay with such a peptide or a co-
CC immunoprecipitation from vertebrate or invertebrate cell lysates or a
CC mammalian or yeast two hybrid assay. The polypeptide and polynucleotide

CC are useful for treating disorders of cell fate, which involves
CC administering therapeutic compounds chosen from invertebrate and
CC vertebrate lgs protein homologues or fragments, antibodies, antibody
CC fragments, lgs antisense DNA, lgs antisense RNA, lgs double-stranded RNA,
CC small peptides or chemical and natural compounds being capable of
CC interfering with lgs function, synthesis and degradation. The disorders
CC are related to cell differentiation or cell proliferation. The compound
CC is administered to treat a cancerous condition by preventing progression
CC from a pre-neoplastic or non-malignant condition to a neoplastic or
CC malignant state. The cancerous condition is characterised by over-
CC stimulation of the wnt pathway and is medulloblastoma or cancer of the
CC colon, breast, head and neck, brain, thyroid or skin. The therapeutic
CC compound may also be administered to a blood disease to promote tissue
CC regeneration and repair. This sequence represents a Drosophila lgs
CC peptide fragment of the invention.
XX
SQ Sequence 112 AA;
Alignment Scores:
Pred. No.: 9.01e-40 Length: 112
Score: 570.00 Matches: 112
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.74% Indels: 0
DB: 8 Gaps: 0
US-10-664-859-1 (1-6909) x ADJ71898 (1-112)
QY 4029 TCTGCGTACGTACCAATAGCCACTCAATCGCCCAATCCCTCGAGTCCAAACAATCTATCT 4088
Db 1 SerAlaSerValProIleAlaThrGlnSerProAsnProSerSerProAsnAsnLeuSer 20
QY 4089 CTCCCGTACCCGGGACAAACCGGACGAGTCATGGGATTGCCGACCAACTCTCTTAGCATG 4148
Db 21 LeuProSerProArgThrThrAlaAlaValMetGlyLeuProThraAsnSerProSerMet 40
QY 4149 GATGAACAGGATCATTTATCTGGATCTGTTCCGCAAGCTAATACTTCGACGGTTGAGGCA 4208
Db 41 AspGlyThrGlySerLeuSerGlySerValProGlnAlaAsnThrSerThrValGlnAla 60
QY 4209 GGCACAACAACAGTGTCTTCAGCAACAAGAACTGTTTTCAGGACAGACACCCCATCGCCG 4268
Db 61 GlyThrThrThrValLeuSerAlaAsnLysAsnCysPheGlnAlaAspThrProSerPro 80
QY 4269 TCAATCAAAATCGTAGTAGAATATACCGGATCGTCAGCGTTCTTACGCATACTTAAGC 4328
Db 81 SerAsnGlnAsnArgSerArgAsnThrGlySerSerSerValLeuThrHisAsnLeuSer 100
QY 4329 AGCAACCCCAAGTACCCCTTATCTCATCTATCCCCA 4364
Db 101 SerAsnProSerThrProLeuSerHisLeuSerPro 112
RESULT 7
ABB11808
ID ABB11808 standard; peptide; 1435 AA.
XX
AC ABB11808;
XX
DT 11-JAN-2002 (first entry)
XX
DE Human BCL9 homologue, SEQ ID NO:2178.
XX
KM Human; cytokine; cell proliferation; cell differentiation; growth factor;
KM haematopoiesis regulation; tissue growth; immunomodulator; activin;
KM inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
KM proliferation; metastasis; cancer; tumour; haematopoietic disorder;
KM myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
KM chronic inflammatory condition; proliferative retinopathy;
KM atherosclerosis; coronary heart disease; arterial ischaemia;
KM bone disorder; osteoporosis; vascular growth disorder;
KM tissue regeneration; wound healing; infection; immune disorder;
KM cell culture; drug screening; gene therapy; antiinflammatory;
KM antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;

| | | | |
|----|------|--|------|
| Db | 360 | GlySerAsnSerSerSerAlaAspProLysAlaProProProProProValSerSergly | 379 |
| QY | 3207 | AGGATCCCTGGCCTGACAACTGAAAACACATTTGTATCACTGCAAGAGTTAAGTTCCA | 3266 |
| Db | 380 | GluproThreLeu---GlyGluAsn-----Pro | 388 |
| QY | 3267 | GACGAAAACCTTACACCAACAACGGGCAACATCGGGAGAAACAGTTGGCAAAAATATAA | 3326 |
| Db | 389 | Asp---GlyLeuSerGlnGlnGlnLeuGluHisArgGluArgSerLeuGlnThrLeuArg | 407 |
| QY | 3327 | AAATGAATCAATTCTTTTCTGAAATGAGAAATTCAGTAGAGCTAATGTAAGCTCA | 3386 |
| Db | 408 | AspIleGlnArgMetLeuPheProAspGluLysGluPheThrGlyAlaGln----- | 424 |
| QY | 3387 | CAGATAACAAAAATTCAGAGAGATTTAATGATGGGATGTGGGTGGCGAGCGGACT | 3446 |
| Db | 425 | -----SerglyGlyPro | 428 |
| QY | 3447 | ATTATAATCCGACGATGGACAACACTGCATATGCCAGTTAACGCCAATCGAGCTTTA | 3506 |
| Db | 429 | GlnGlnAsnPro----- | 432 |
| QY | 3507 | TCCGCGACAAGTTCAGACTTTCGGAAGATGTAATGCATCCAGGGATGTTATATCAGAT | 3566 |
| Db | 432 | ----- | 432 |
| QY | 3567 | ATGGGTCCGTTAATAGGATGTAATATAATCAAAAAACAGTGTGCAATGTGATTCGA | 3626 |
| Db | 433 | ---GlyValLeuAspGlyProGlnLysLysProGlnGlyProIleGlnAla----- | 448 |
| QY | 3627 | GTAGGTGTGTCACTGGAACAATCGAGCTGAGTAAATGTCATATGCATTTGCTCAAGC | 3686 |
| Db | 449 | -----MetMetAlaGlnSer | 453 |
| QY | 3687 | TCCGCGCCCCGAATGGCAATATGATGGAGCTGTACGATATGCTAGCCTCGTTGGC | 3746 |
| Db | 454 | GlnSerLeuGlyLysGly-----ProGlyProArgThrAspValGlyAlaProPheGly | 471 |
| QY | 3747 | AACACAAGCTGCAACGTCATCGGAACGGCCCCAGAT--ATGCTAAGGAAGTTTAAAT | 3803 |
| Db | 472 | ProGlnGlyHisArgAspValProPheSerProAspGluMetValProProSerMetAsn | 491 |
| QY | 3804 | CAAGTAGCCGAACC-----CATTCACATCAAGGGGAGTTGCTCAATGGAG | 3851 |
| Db | 492 | SerGlnSerGlyThrIleGlyProAspHisLeuAspHisMetThrProGlnGlnIleAla | 511 |
| QY | 3852 | TGGTCGAAGATTCAACATCAATTTTTCGAAGAACGCTCAAGGGGGGCAAGCCCAAGCAA | 3911 |
| Db | 512 | TrpLeuLysLeuGlnGlnGluPheTyrGluLysArgArg---LysProGlnGln | 529 |
| QY | 3912 | GTCACGTGAACGTGATACCAAA-----CAGCAAAAC | 3944 |
| Db | 530 | Val-----ValValGlnGlnCysSerSerLeuGlnAspMetMetValHisGlnHisGly | 546 |
| QY | 3945 | CCTTCTGATCTGGTGAAACTCGTTAAACACCAAGGTGCAACCCCTGCAGAGTCCACT | 4004 |
| Db | 547 | ProArgGly-----ValValArgGlyProPro | 555 |
| QY | 4005 | CCTCCTTACCAC----- | 4016 |
| Db | 556 | ProProTyrGlnMetThrProSerGlnGlyTyrAlaProGlyGlyThrGluProPheSer | 575 |
| QY | 4017 | ---TCCATCCAGAGATCTGGCTCAGTACCAATAGCCACTCAATCGCCCAATCCTCGAGT | 4073 |
| Db | 576 | AspGlyIleAsnMetProHisSerLeuProProArgGlyMetAlaProHisProAsnMet | 595 |
| QY | 4074 | CCAAACAATCTA----- | 4085 |
| Db | 596 | ProGlySerGlnMetArgLeuProGlyPheAlaGlyMetIleAsnSerGluMetGlnGly | 615 |
| QY | 4086 | ---TCTCTCCCGTACCGCGGACAACCGCAGCAGTCAATGGAGTTG-----CCGACCAAC | 4136 |

| | | | | |
|---|---|------|--|------|
| D | b | 616 | ProAsnValProAsnProAlaSerArgProGlyLeuSerGlyValSerTrpProAspAsp | 635 |
| Q | Y | 4137 | TCTCCTAGCATG-----GATGGAACAGATCATTATCTGA--- | 4172 |
| | | | ::: | |
| D | b | 636 | ValProLysIleProAspGlyArgAsnPheProProGlyGlnGlyIlePheSerGlyPro | 655 |
| Q | Y | 4173 | ----- | |
| D | b | 656 | GlyArgGlyGluArgPheProAsnProGlnGlyLeuSerGluGluMetPheGlnGln | 675 |
| Q | Y | 4191 | ACTTCGACGGTTTCAGGCGAGGACACACA-----ACAGTCTCTCAGCAAAACAAGACTGT | 4244 |
| | | | ::: ::: | |
| D | b | 676 | LeuAlaGluLysGlnLeuGlyLeuProProGlyMetAlaMetGluGlyIleArgProSer | 695 |
| Q | Y | 4245 | TTTCAGGCAGACACCCCATCGCCGTCAAATCAAAATCGTAGTAGAAATACCGGATCGTCA | 4304 |
| | | | :::>::: ::: | |
| D | b | 696 | MetGluMetAsnArgMetIleProGlySerGln--ArgHisMetGluProGlyAsnAsn | 714 |
| Q | Y | 4305 | AGCGTCTTACGCACTTAAGCAGCAACCCCAAGTACCCCTTATCTCATCTATCCCA | 4364 |
| | | | :::>::: ::: | |
| D | b | 715 | ProIlePheProArgIleProValGluGlyProLeuSerProSerArgGlyAspPhePro | 734 |
| Q | Y | 4365 | AAG-----GAATTGAGTCTTTCGTCAGTCTCT | 4394 |
| | | | ::: | |
| D | b | 735 | LysGlyIleProProGlnMetGlyProGlyArgGluLeuGlu--PheGlyMetValPro | 753 |
| Q | Y | 4395 | GCTGTATGTTATATTGTTTAAATTTTAAAGACAATCAATATATGAATTGCGTTAAT | 4454 |
| | | | ::: ::: | |
| D | b | 754 | SerGlyMet-----LysGlyAspValAsnLeuAsn--ValAsn | 765 |
| Q | Y | 4455 | AATAAGTTATATATACATAACTCGGAATTTGATGAAAAAATCAGGAATAGAAAAAT | 4514 |
| | | | ::: ::: | |
| D | b | 766 | -----MetGlySerAsnSerGlnMetIleProGlnLysMetArg | 778 |
| Q | Y | 4515 | AAATTATTTCGGACCG----- | 4532 |
| | | | ::: ::: | |
| D | b | 779 | GluAlaGlyAlaGlyProGluGluMetLeuLysLeuArgProGlyGlySerAspMetLeu | 798 |
| Q | Y | 4532 | ----- | 4532 |
| D | b | 799 | ProAlaGlnGlnLysMetValProLeuProPheGlyGluHisProGlnGlnLutyrGly | 818 |
| Q | Y | 4533 | -----CCCATCCATTCTTG-----AATCCAATTTCTGAGTATGTTAGA | 4574 |
| | | | ::: | |
| D | b | 819 | MetGlyProArgProPheLeuPrometSerGlnGlyProGlySerAsnSerGlyLeuArg | 838 |
| Q | Y | 4575 | GATAA-----TCTACTATTAAATTAACAAGAAAT | 4606 |
| | | | :::>::: ::: | |
| D | b | 839 | AsnLeuArgGluProIleGlyProAspGlnArgThrAsnSerArgLeuSerHisMetPro | 858 |
| Q | Y | 4607 | TCATATCCGTTAATTGAAAAATCACTATTGTTAATAAGAAATTAAAAATATGTTATTAT | 4666 |
| | | | | |
| D | b | 859 | ProLeuProLeu----- | 862 |
| Q | Y | 4667 | AATATTCTACAGGTGATAACATGAAAAAGTAGCGGACCAAGCCACAGGGTCAGCGGTCA | 4726 |
| | | | :::>::: ::: | |
| D | b | 863 | -----AsnProSerSerAsnProThrSerLeuAsnThrAlaProProValGlnArg--- | 879 |
| Q | Y | 4727 | CCAGTAATAGTCTAATAGAGGCAATAAAGATGTACGATTGCTGCATCC----- | 4777 |
| | | | ::: | |
| D | b | 880 | -----GlyLeuGlyArgLysProLeuAspIleSerValAlaGlySerGlnValHis | 896 |
| Q | Y | 4778 | AGTCTGTTTAAACCCGCATCCACATATGCAAGAACAATTCAAATTCAGCATTAACGCC | 4837 |
| | | | ::: | |
| D | b | 897 | SerProGlyIleAsnPro-----LeuLysSerProThrMetHisGlnValGlnSer | 913 |
| Q | Y | 4838 | TATAAATGGGCTCTACCAATATACAGATGAGGTAAATATTTAATATTTTAAACG | 4897 |
| | | | ::: ::: | |
| D | b | 914 | PrometLeuGlySerProSerGlyAsnLeu----- | 923 |
| Q | Y | 4898 | TTTTGTGTTAATTATCTTCTTTTCAGCGGTCAAGCATCAGCGCAAGGTGATCCGTAC | 4957 |
| | | | ::: | |
| D | b | 924 | -----LysSerProGlnThrProSer | 930 |

```

QY 4958 AATTAGTCGCGC-----GCTCCGATAATATTCGCTAAATCCCAATA 4999
Db      ||| |||
Db 931 GlnLeuAlaGlyMetLeuAlaGlyProAlaAlaAlaSerIleYSerProProVal 950
QY 5000 GTGCAATCGCGCGC-----
Db      ||| ||| ||| |||
Db 951 LeuGly-SerAlaAlaAlaSerProValHisLeuYSerProSerLeuProAlaProse 970
QY 5016 -----ACCAACAAGATGACCCCAAACTTCGATCCCAATCTTCTTTGGACAAA 5065
Db      ||| ||| ||| ||| |||
Db 970 rProGlyThrThrSerSerProGluProProLeuGlnSer-----ProG1 985
QY 5066 TGTCCCAACAATAACAAGTTCGCTGTCAGCATGGGTAGTCCAGCCGGAACGTGTGTA 5125
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 985 YIleProProAsnHisIlySAlaProLeu-ThrMetAlaSerProAlaMetLeuGlyAsnV 1005
QY 5126 TGACGATGATGGGGGTCCG-----GGACCGTCCGACATCAATATTGAGC 5170
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1005 AlGluSerGlyGlyProProProProThraIaSerGlnProAlaSerValAsnIleProG 1025
QY 5171 ATGAATAATTCCGGACTAGATGATCA-----GGAATAGATACCAATAATCAAAATA 5224
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1025 IySerLeuProSerSerThrProTyThrMetProProGluProThrLeuSerGlnAsnP 1045
QY 5225 ACTGTCATTCATGAATGTCGTATGAATCAATGGGTCCCGCAATGTGATCTCT---- 5280
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1045 roLeuSerIleMetMetSerArgMetSerIlySPhelaIaMetProSerSerAsnProGlyT 1065
QY 5281 -----AAATGTCCGTACAGAGCGGTCCAATATGACCCGCTGGCTTTA 5323
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1065 yTrAsnHisAspAlaIleIySThrValAlaSerSerAspAspAspSerProProAlaArgS 1085
QY 5324 ATCCTAATTCCTCCCAATGGTGGATTAAAGAGAATTCATAGGGTCTGCTGTGCTCAG 5383
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1085 erProAsnLeuProSer-----MetAsnAsnMetProGlyMetGlyIleAsnThrG 1102
QY 5384 CAAAC----- 5388
Db      ||| |||
Db 1102 InAsnProArgIleSerGlyProAsnProValValPrometProThrLeuSerPrometG 1122
QY 5389 -----TCTTCAACTTTCAAGGGTT- 5409
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1122 IyMetThrGlnProLeuSerHisSerAsnGlnMetProSerProAsnAlaValGlyProA 1142
QY 5410 -----GTTCCACCTGTGTCGAGAATGATGGTGCATGCCAGTC--A 5449
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1142 snIleProProHisGlyValPrometGlyProGlyLeuMetSerHisAsnProIleMetG 1162
QY 5450 ATTTGGTTCGAATTTCAATCCGAATATT--CAGGTAAGGCGAGTACCCCAAC--- 5502
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1162 IyHisGlySerGlnGluProPrometValProGlnGlyArgMetGlyPheProGlnGlyP 1182
QY 5503 -----ACCATACATATCATGCCAGTAAGGCGACAGAACGCCAACACATTAACAACATG 5557
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1182 heProProValGlnSerProProGlnGlnValProPheProHisAsnGlyProSerGlyG 1202
QY 5558 GAGCTAATAATGTCGAATGCCACCTAGTCTGGAATTTTTCAGAGGTACGCTAACCCCTC 5617
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1202 IyGlnGlySer-----PheProGlyGlyMetGlyPhe-----ProG 1214
QY 5618 AAATGGTGTGTAGGCAATGGTCCGCATATATGCCACCATCAGCCAGCAGCGTACTC 5677
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1214 IyGlnGlyProLeuGlyArgProSerAsnLeu---ProGlnSerSerAlaAspAlaAla 1233
QY 5678 CTGGAATGCCAGATTGATGGCGGAGCCAGAGCCGAGGTATGCTAATGAATCT---- 5733
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1233 euCySlySProGlyGlyProGlyGlyProAspSerPheThrValLeuGlyAsnSerMetP 1253
QY 5733 ----- 5733
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1253 roSerValPheThrAspProAspLeuGlnGluValIleArgProGlyAlaThrGlyIleP 1273

```

```

QY 5733 ----- 5733
Db 1273 roGluPheAspLeuSerArgIleIleProSerGluIlySProSerGlnThrLeuGlnTyP 1293
QY 5734 -----TCCGAGAGCAACACCAGACAAGATCACAACAACATCTGGGGCAAGCAATGTA 5788
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1293 heProArgGlyGluValProGlyArgIySglnProGlnGlyProGlyPro-----GlyP 1311
QY 5789 TTAATCTTTCAGAAATTCGAATGCAATGTCTATTGTTGACGAAGAGGT----- 5838
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1311 heSerHisMetGln-----GlyMetMetGlyGlnGlnAlaProArgMetG 1326
QY 5839 -----GATTAACCGCGCATGACGGATCAATGAATATTTGTTCAACCATCTATGATAAGG 5893
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1326 IyLeuAlaLeuProGlyMetGlyGlyProGlyProValGlyThrProAspIleProLeuG 1346
QY 5894 GCATGCGTCCA-----CATGCCATGCGGCCAAAT-----GTAATGGGTG 5932
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1346 IyThrAlaProSerMetProGlyHisAsnPrometArgProProAlaPheLeuGlnGln 1366
QY 5933 CCGCGATGCCACCCGTTAACAGGCAATTCAGTTTGACAGTCAATCGGATGTATTGACT 5992
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1366 IyMetMetGlyProHisHisArgMetMetSerProAlaGlnSerThr-----M 1382
QY 5993 GTGTCCGGGATCCGTCATCATTTTTCATTAACGCTTCTCTGCAACAGCGCTGACACACA 6052
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1382 erProGlyGlnProThrLeuMetSerAsnProAlaAlaAla----- 1395
QY 6053 TGTTTGATCAGCACACAGGCCAATCAGCCTAAGACACACAATAAAGACATACCTA 6112
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1396 -----ValGlyMetIlePro- 1400
QY 6113 GTGAATGTGTCAAAACCAATCGGACTTGCAGTGACAGACAGGAGATCCAACTGCATG 6172
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1401 --GlyIySAspArgGlyProAlaGlyLeuTyThrHisProGlyProValGlySerProG 1420
QY 6173 GGCAAGACATCGCGCAGGTCAGTCTTAAATGAGCCTACTATAATAAT 6222
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1420 IyMetMetMetSerMet--GlnGlyMetMetGlyProAsnArgThrSer 1435

RESULT 8
ADQ18945
ID ADQ18945 standard; protein; 1394 AA.
XX
AC ADQ18945;
XX
DT 26-AUG-2004 (first entry)
XX
DE Human soft tissue sarcoma-upregulated protein - SEQ ID 1764.
XX
KW soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human.
XX
OS Homo sapiens.
XX
PN W02004048938-A2.
XX
PD 10-JUN-2004.
XX
PF 26-NOV-2003; 2003WO-US038193.
XX
PR 26-NOV-2002; 2002US-0429739P.
XX
PA (PROT-) PROTEIN DESIGN LABS INC.
XX
PI Aziz N, Ginsburg WM, Zlotnik A;
XX
DR WPI; 2004-441208/41.
XX
PT Early detection of soft tissue sarcoma comprises determining expression
PT of a gene in a first soft tissue sample and a normal soft tissue sample
PT and comparing the gene expression, also useful in treating soft tissue
PT sarcoma.
XX

```

PS Example 2; SEQ ID NO 1764; 210pp; English.

XX The invention relates to a novel method for detecting soft tissue sarcoma
CC which comprises obtaining a first soft tissue sample from an individual
CC and a normal soft tissue sample from the same or different individual,
CC determining the expression of a gene in both samples and comparing the
CC expression of the gene in both soft tissue samples, where a higher level
CC of protein expression in the first soft tissue sample indicates the
CC presence of soft tissue sarcoma. The method of the invention has
CC cytostatic applications and may be useful for detecting soft tissue
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC acid sequences may be useful in diagnostic and screening applications.
CC The current sequence is that of a human soft tissue sarcoma-upregulated
CC protein of the invention. The current sequence is not shown within the
CC specification per se but was submitted in CD format by the inventor.

XX Sequence 1394 AA;

Alignment Scores:

| | | | |
|------------------------|----------|---------------|------|
| Pred. No.: | 3.12e-17 | Length: | 1394 |
| Score: | 314.00 | Matches: | 306 |
| Percent Similarity: | 30.12% | Conservative: | 172 |
| Best Local Similarity: | 19.28% | Mismatches: | 509 |
| Query Match: | 2.61% | Indels: | 601 |
| DB: | 8 | Gaps: | 70 |

US-10-664-859-1 (1-6909) x ADQ18945 (1-1394)

QY 2397 AAAATGGAAGA--ATTTCAAACGACAGTACCAGAAAAAAGATCGTCTTGACA 2453
Db 98 LysArgGluArgSerIleSerAlaAspSerPheAspGlnArgAspProGlyThr----- 115
QY 2454 ATGAATAATGACGAAATGAGCATGAGAGCGTCAATCAGTTGAATCCCGATTATCAAT 2513
Db 116 ---ProAsnAspAspSerAspIleLysGluCysAsnSerAlaAspHisIleLysSerGln 134
QY 2514 GAATCTTAAATAATCTCTCAATTTGAGCATATAGTAAGCGAGTAGACCAATACCC 2573
Db 135 AspSerGlnHisThrProHisSerMetThr----- 144
QY 2574 GGAATCGAGTTGAGCGGGGACGGGAATTTATGACTGCCACGCCAATGGAATCTCC 2633
Db 145 -----ProSerAsnAlaThrAlaProArgSerSerThrProSer 157
QY 2634 TCGGAGTACGTAATGTTGGATTACATGCAACGCAAAATCAC-----ATA 2681
Db 158 HisGlnGlnThrThrAlaThrGluProThrProAlaGlnLysThrProAlaLysValVal 177
QY 2682 TTCGTGTTTCACTCAGCTGGCCACAAGGGCGCAATCAGTTTAAGCGTCAATTT 2741
Db 178 TyrValPheSerThrGluMetAlaAsnLysAlaIleGluAlaValLeuLysGlnVal 197
QY 2742 CAAACTATTATGCGTATCACTGCACTCAGCCTGCTACAAAAAGCTTCTGGAAGACTTT 2801
Db 198 GluThrIleValSerPheHis----- 204
QY 2802 TTTATGAAAAACCCCTTAAGATTAAAGTTACAGGGCACAATTCGTCGTATGCCA 2861
Db 205 ---IleGlnAsn---IleSerAsnAsnLysThrGluArgSerThrAla-----Pro 219
QY 2862 TGGATAGGCATGGGCGAGGTTGGACTAATCTCTTAATCCTGTAGCCAAAAATAACAAA 2921
Db 220 LeuAsnThrGlnIleSerAlaLeuArgAsnAspProLysProLeu-----Pro 235
QY 2922 CAGCAGCCACATACAAAGACCGTAGGCTATTGAAACCCCAATTCAATCAACATGAAAAC 2981
Db 236 GlnGlnProProValProAla-----AsnGlnAspGlnAsn 247
QY 2982 AGCAACGTAAGTACTGTAAAGCGCGCTAGCAACTCTTT----- 3020
Db 248 SerSerGlnAsnThrArgLeuGlnProThrProIleProAlaProAlaProLysPro 267
QY 3021 -----GTGACCAGTCTGATCTATGGGCAACGAAACTGAATTGATG 3062

Db 268 AlaAlaProProArgProLeuAspArgGluSerPro---GlyValGluAsnLysLeuIle 286
QY 3063 TGCTGGAGAGCGGATCTCTCAACACACCAGT-----AGTTCGACAAAAC 3107
Db 287 ProSerValGlySerProAlaSerSerThrProLeuProProAspGlyThrGlyProAsn 306
QY 3108 TCACGAATCATGTAGACAGTATCAGTATCCAGCGAGTCACAGGCAATAAAGATTAAGTCTG 3167
Db 307 SerThrProAsnAsnArgAlaValThrProValSerGlnGlySerAsnSerSerAla 326
QY 3168 GAAGCAGCTGGCGTTGATTGGACAGGTCAAAAAGAGCGATCTGGCTGACAAC 3227
Db 327 AspProLysAlaProProProProProValSerSerGlyGluProProThrLeu---Gly 345
QY 3228 GAAACAACATGTATCACTGCAAGAGTTAAGTTCAGACGAAAACTTACACACACA 3287
Db 346 GluAsn-----ProAsp---GlyLeuSerGlnGlu 354
QY 3288 CAGCGGCAACATCCGGGAAGACAGTTGGCAAAAATAAAAAATGAATCAATTTCTTTT 3347
Db 355 GlnLeuGlnHisArgGluArgSerLeuGlnThrLeuArgAspIleGlnArgMetLeuPhe 374
QY 3348 CCTGAAATGAGAAATTCACTAGAGCTTAATGTAAAGCTCACAGATTAACAAAATTCCAGGA 3407
Db 375 ProAspGluLysGluPheThrGlyAlaGln----- 384
QY 3408 GATTTAATGATGGGATGTGCGGTGGCGGAGCGGATCTATTATTAATCCGACGATCGGA 3467
Db 385 -----SerGlyGlyProGlnGlnAsnPro----- 392
QY 3468 CAACTGCATATGCCAGGTAACGCCAAATCGAGCTTTATCGGCGACAAAGTTCAGACTT 3527
Db 392 ----- 392
QY 3528 TCGGAAGATGTAATGCATCCAGGGATGTATATCAGATATGCGTCCGTAATAGATGT 3587
Db 393 -----GlyValLeuAspGlyPro 398
QY 3588 AATAATATCAAAAAACCAAGTGTGCAATGTGATCTGAGTAGGTGTTCACCTGAACA 3647
Db 399 GlnLysLysProGluGlyProIleGlnAla----- 408
QY 3648 ACTGCAGCTGGAATAATGTCAATATGCAATGCTCAAGCTCCGGCGCCGCAATGCCAAT 3707
Db 409 -----MetMetAlaGlnSerGlnSerLeuGlyLysGly--- 419
QY 3708 ATGATGGAGCTCTACGGATATGCTAGCTCTGTTGGCAACACAAAGTGCACGTCATC 3767
Db 420 ---ProGlyProArgThrAspValGlyAlaProPheGlyProGlnGlyHisArgAspVal 438
QY 3768 GGAACGCCCCAGAT--ATGCTAAGGAAGTTTAAATCAAGATAGCCGAACC----- 3818
Db 439 ProPheSerProAspGluMetValProProSerMetAsnSerGlnSerGlyThrIleGly 458
QY 3819 -----CATTCAATCAAGGGGAGTTGCTCAAAATGAGTGTGCAAGATTCACATCAA 3872
Db 459 ProAspHisLeuAspHisMetThrProGluGlnIleAlaThrLeuLysLeuGlnGlnGlu 478
QY 3873 TTTTTCGAAGACGCTCAAGGGGGGCAAGCCAGACAAGTCACTGGAAGTGTAGTACCA 3932
Db 479 PheTyrGluGluLysArgArg-----LysProGluGlnVal-----ValValGln 493
QY 3933 CAA-----CAGCAAAACCCCTTGTGATCTGTGTGGAAC 3965
Db 494 GlnCysSerLeuGlnAspMetMetValHisGlnHisGlyProArgGly----- 509
QY 3966 TCGTTAAACAACAGGTGCGACCCCTGCAAGTCCACTCTCTTACCA----- 4016
Db 510 -----ValValArgGlyProProProProTyrGlnMetThrPro 522
QY 4017 -----TCCATCCAGAGATCTGCG 4034

Db 523 SerGluGlyTyrPheAlaProGlyGlyThrGluProPheSerAspGlyIleAsnMetProHis 542
Qy 4035 TCAGTACCAATAGCCACTCAATCGCCCAATCCCTCGAGTCCAAACAATCTA----- 4085
|||:::|
Db 543 SerLeuProProArgGlyMetAlaProHisProAsnMetProGlySerGlnMetArgLeu 562
Qy 4086 -----TCTCTCCCGTCACCGCGG 4103
:::|
Db 563 ProGlyPheAlaGlyMetIleAsnSerGlnMetGluGlyProAsnValProAsnProAla 582
Qy 4104 ACAACCGCAGCAGTCATGGATTG-----CCGACCAACTCTCTAGCATG----- 4148
:::|
Db 583 SerArgProGlyLeuSerGlyValSerTrpProAspAspValProLysIleProAspGly 602
Qy 4149 -----GATGGAACAGATCATTTATCTGA----- 4172
|||
Db 603 ArgAsnProProGlyArgGlyIlePheSerGlyProGlyArgGlyGluArgPhePro 622
Qy 4173 -----TCTGTTCCGCAAGCTAATACTTCGACGGTTCAGCAGGC 4211
:::|
Db 623 AsnProGlnGlyLeuSerGluGluMetPheGlnGlnGlnLeuAlaGlyLysGlnLeuGly 642
Qy 4212 ACAACA-----ACAGTCTCTCAGCAACAACAACACTGTTTTCAGCGCACACCCCATCG 4265
:::|
Db 643 LeuProProGlyMetAlaMetGluGlyIleArgProSerMetGluMetAsnArgMetIle 662
Qy 4266 CCGTCAAAATCAAAATCGTAGTAGAATAACCGGATCGTCAGCGTTCCTTAACGATTA 4325
|||:::|
Db 663 ProGlySerGln--ArgHisMetGluProGlyAsnAsnProIlePheProArgIlePro 681
Qy 4326 AGCAGCAACCAAGTACCCCTTATCTCATCTATCCCAAG----- 4367
|||
Db 682 ValGluGlyProLeuSerProSerArgGlyAspPheProLysGlyIleProProGlnMet 701
Qy 4368 -----GAATTTGAGTCTTTCGGTCAGTCTCTGCTGATGTATATTGTTT 4415
|||
Db 702 GlyProGlyArgGluLeuGlu--PheGlyMetValProSerGlyMet----- 716
Qy 4416 AATTTTAAAGACAATAATGAATGCGTTAATAATAAGTTATATATATACATAA 4475
|||
Db 717 -----LysGlyAspValAsnLeuAsn--ValAsn----- 725
Qy 4476 CTCGAAATTTGATAGAAAATAATCAGGAATAGAAAAATAATTATTTCCGACCG-- 4532
:::|
Db 726 MetGlySerAsnSerGlnMetIleProGlnLysMetArgGluAlaGlyAlaGlyProGlu 745
Qy 4532 ----- 4532
Db 746 GluMetLeuLysLeuArgProGlyGlySerAspMetLeuProAlaGlnGlnLysMetVal 765
Qy 4533 -----CCGATCCATTCTTG 4547
|||
Db 766 ProLeuProPheGlyGluHisProGlnGlnGluTyrGlyMetGlyProArgProPheLeu 785
Qy 4548 -----AATCCAATTTCTGAGTGAATTGTAGAGATA----- 4579
|||
Db 786 PrometSerGlnGlyProGlySerAsnSerGlyLeuArgAsnLeuArgGluProIleGly 805
Qy 4580 -----TCTACTATTAAATTAACAACGAAAAATTCATATCCGTTAATTGAAAT 4627
:::|
Db 806 ProAspGlnArgThrAsnSerArgLeuSerHisMetProProLeuProLeu----- 822
|||
Qy 4628 CACTATTGTTTAATAAGAAATTAAAAATATGTTATTATATATTCTACAGGTGATAAC 4687
:::|
Db 823 -----AsnProSerSerAsn 827
Qy 4688 ATGAAAGTAGCGCACCAAGCCACAGGTCACCGTCCACCAGTAATAAGTCTAATAGAG 4747
|||
Db 828 ProThrSerLeuAsnThrAlaProProValGlnArg-----GlyLeuGlyArg 843
Qy 4748 GCAAAATAAGATGATGATTGCTGCATCC-----AGTCCGTGTTTAAACCCGCAT 4798
|||
Db 844 LysProLeuAspIleSerValAlaGlySerGlnValHisSerProGlyIleAsnPro--- 862

Qy 4799 CCACATATGCAAAGCAATTCAAATTCAGCATTAACGCCCTATAAATGGGCTTACCAAT 4858
:::|
Db 863 -----LeuLysSerProThrMetHisGlnValGlnSerProMetLeuGlySerProSer 880
Qy 4859 ATACAGATGAGGTAAATATATTAAATATTTTATTAAAGTTTGTGTATTATCTTC 4918
:::|
Db 881 GlyAsnLeu----- 883
Qy 4919 TTTTTCAGCGTCAAGCATCAGCGCAAGGTGATCCGTACAATTAGTCGGC----- 4969
|||
Db 884 -----LysSerProGlnThrProSerGlnLeuAlaGlyMetLeuAla 897
Qy 4970 -----GCTCCGATATATTCGGCTAAATCCCAATAGTGGCAATCGCGGCC----- 5015
|||
Db 898 GlyProAlaAlaAlaAlaSerIleLysSerProProValLeuGly-SerAlaAlaAlaSer 917
Qy 5016 -----ACCAACAAGA 5026
|||
Db 917 rProValHisLeuLysSerProSerLeuProAlaProSerProGlyTyrThrSerSerPr 937
Qy 5027 TGACCCAAAACCTTCGATCCAAATCTTCTTTGGCACAAATGTCCCAACAATAACAAGTT 5086
|||
Db 937 OlyProProLeuGlnSer-----ProGlyIleProProAsnHisLysAl 952
Qy 5087 GCGTTCACGATGGGTAGTCCAGCCGGAACCTGCTGATGACGATGATGGGGTCCG- 5145
:::|
Db 952 aProLeu-ThrMetAlaSerProAlaMetLeuGlyAsnValGluSerGlyGlyProProP 972
Qy 5146 -----GGACCGTCCGACATCAATATTGAG----- 5169
|||
Db 972 roProThrAlaSerGlnProAlaSerValAsnIleProGlySerLeuProSerSerThrP 992
Qy 5169 ----- 5169
Db 992 roTyrThrMetProProGluProThrLeuSerGlnAsnProLeuSerIleMetMetSera 1012
Qy 5170 -----CATGAATAATTTCGGGAC 5188
|||
Db 1012 rgMetSerLysPheAlaMetProSerSerThrProLeuTyrHisAspAlaIleLysThrV 1032
Qy 5189 TAGATGATCAGGAATAGATACC-----ATAAATCAAAATGACTGTCAATCAATGAATG 5242
:::|
Db 1032 alAlaSerSerAspAspAspSerProProAlaArgSerProAsnLeuProSerMetAsn- 1051
Qy 5243 TCGTAATGAATCAATGGGTCCCGCAATGCTGAATCCTAAATGTCGCTAGCAGCGGCTC 5302
|||
Db 1052 --AsnMetProGlyMetGlyIleAsnThrGlnAsnProArgIle-----SerGlyP 1068
Qy 5303 CAAAT----- 5307
|||
Db 1068 roAsnProValValProMetProThrLeuSerProMetGlyMetThrGlnProLeuSerH 1088
Qy 5308 -----GGACCGCTGGCTTTAATCCTTAATTCCCAATGGTGATTA 5350
|||
Db 1088 isSerAsnGlnMetProSerProAsnAlaValGlyProAsnIleProProHisGlyVal- 1107
Qy 5351 GAGAGAAATTCATAGGCTGGCTGTGGCTCAGCAAACTCTTCAACTTTCAGGGGTTG 5410
:::|
Db 1108 -----PrometGlyProGlyLeuMetSerHisAsnProIleMetGlyHisGlySerG 1125
Qy 5411 TTCACCTGTGTCCAGATGATGGGTGGAATGCCAGTCAATTTGGTTCGAATTTCAATC 5470
|||
Db 1125 InGluProProMetValProGlnGlyArgMet-----GlyPheProGlnGlyPheProP 1143
Qy 5471 CGAATATTCAAGTAAAGGCGAGTACCCCAACAACCATACATACATGCCAGTAAGGCAC 5530
|||
Db 1143 ro-----ValGlnSerProProGlnGlnValP 1152
Qy 5531 AGAACGCCAACAACAATAACAACAATGGAGCTAATAATGTGCGAATGCCACCTAGTCTGG 5590
:::|
Db 1152 roPheProHisAsnGlyProSerGlyGlyGlnGlySer-----PheProGlyGlyMetG 1170

QY 5591 AATTTTGACAGAGTACGCTAACCCCTCAATGGCTGTGTAGGCAATGGTCCCAATAT 5650
Db 1170 lyphe-----ProGlyGlnGlyProLeuGlyArgProSerAsnLeu- 1183
QY 5651 GCCCACCATGACCGACGACGGTACTCCTGGAATGCCAGATTGATGCGGACGACGAG 5710
Db 1184 --ProGlnSerSerAlaAspAlaAlaLeuGlySerProGlyGlyProGlyGlyProAsps 1203
QY 5711 CCGAGGATATGCTAATGATTCCT----- 5733
Db 1203 erPhetThrValLeuGlyAsnSerMetProSerValPhetThrAspProAspLeuGlnIuv 1223
QY 5733 ----- 5733
Db 1223 alIleArgProGlyAlaThrGlyIleProGlnIuPheAspLeuSerArgIleIleProSerG 1243
QY 5734 -----TCCGAGAGCAACACAGACAGATCA 5761
Db 1243 lulySProSerGlnThrLeuGlnTyrPheProArgGlyGluValProGlyArgIlySglnP 1263
QY 5762 CAACAATCCTGGGCAAGCAATGTATTAACTTTTCAGAAATGCAATCAAAATGTCTA 5821
Db 1263 roGlnGlyProGlyPro-----GlyPheSerHisMetGln-----GlyM 1276
QY 5822 TTGTTGACGAGAGGGT-----GGAATTACCCGCGCATGACGATCAATGA 5866
Db 1276 etMetGlyGlnGlnAlaProArgMetGlyLeuAlaLeuProGlyMetGlyGlyProGlyP 1296
QY 5867 ATATTGGTCAACCATCTATGATTAAGGGCGCATGCGTCCA-----CATGCCA 5911
Db 1296 roValGlyThrProAspIleProLeuGlyThrAlaProSerMetProGlyHisAsnProm 1316
QY 5912 TCGCGCCAAAT-----GTAATGGGTGCGCGCATGCCACCCGTTAACAGGCAAAATTCAGT 5965
Db 1316 etArgProProAlaPheLeuGlnGlnGlyMetMetGlyProHisHisArgMetSerP 1336
QY 5966 TTGCACAGTCATCGATGTGATTGACTGTGTCCGGGATCCGTCATCATTTTTCACCTAACG 6025
Db 1336 roAlaGlnSerThr-----MetProGlyGlnProThrLeuMetSerAsnProA 1352
QY 6026 CTTCTGCAACAGCGCTGGACCAACACATGTTTGATTCAGCAACAGGCCAATCAGCCTA 6085
Db 1352 laAlaAla----- 1354
QY 6086 AGACACAACACATAAAGACATACCTAGTGAATGTGTCAAAACCAATCCGGACTTGCA 6145
Db 1355 -----ValGlyMetIlePro--GlyIlyAspArgGlyProAlaGlyLeuTyrT 1370
QY 6146 TGGCACAAGGCGCATCTCAACTGCATGGGCAAGACATCGCGAGGTCACTTTTAATTG 6205
Db 1370 hrHisProGlyProValGlySerProGlyMetMetSerMet--GlnGlyMetMetG 1389
QY 6206 GACCTACTAATAATAAT 6222
Db 1389 lypProAsnArgThrSer 1394
RESULT 9
AAB71229
ID AAB71229 standard; protein; 1426 AA.
XX
AC AAB71229;
XX
DT 18-NOV-2002 (first entry)
XX
DE Human legless homologue lgs/bcl9 protein.
XX
KW Legless; human; lgs; Wnt/Wingless signaling pathway; Wnt; Wg;
KW tissue proliferation; tumour; cytosstatic; cellular disorder; colon;
KW blood disorder; cancer; breast; head and neck cancer; brain; thyroid;
KW medulloblastoma; skin cancer; tissue regeneration; tissue repair.
XX
OS Homo sapiens.
XX

PN US2002086986-A1.
XX
PD 04-JUL-2002.
XX
PF 27-JUL-2001; 2001US-00915543.
XX
PR 28-JUL-2000; 2000US-0221502P.
XX
PA (BASL/) BASLER K.
PA (BRUN/) BRUNNER E.
PA (FROE/) FROESCH B.
PA (KRAM/) KRAMPS T.
PA (PETE/) PETER O.
XX
PI Basler K, Brunner E, Froesch B, Kramps T, Peter O;
XX
DR WPI; 2002-635689/68.
XX
DR N-PSDB; AAF88467.
XX
PT Novel polypeptide useful in therapeutic method for treating disorders of
XX cell fate such as cell differentiation or cell proliferation.
XX
PS Example II; Fig 8B; 41pp; English.
XX
CC This invention describes a novel polypeptide sharing one or more
XX homologous amino acid domains with the legless (lgs) protein, a
XX downstream component of the Wnt/Wingless (Wnt/Wg) signaling pathway
XX involved in the formation and maintenance of spatial arrangements and
XX proliferation of tissues during development, and in the formation and
XX growth of many human tumours. The products of the invention have
XX cyrostatic activity and can be used to treat cellular disorders, blood
XX disorders and cancers caused by over-stimulation of the Wnt pathway,
XX where the cancerous condition is colon, breast, head and neck, brain,
XX thyroid, medulloblastoma or skin cancer. The product could also be used
XX to promote tissue regeneration and repair. This sequence represents the
XX human legless (lgs) protein homologue lgs/bcl9 described in the
XX disclosure of the invention
XX
SQ Sequence 1426 AA;
Alignment Scores:
Pred. No.: 4.69e-17 Length: 1426
Score: 312.00 Matches: 313
Percent Similarity: 30.34% Conservative: 177
Best Local Similarity: 19.38% Mismatches: 519
Query Match: 2.59% Indels: 607
DB: Gaps: 73
US-10-664-859-1 (1-6909) x AAB71229 (1-1426)
QY 2397 AAAATTGAAGA--ATTCAACGACAGTACCACGGAAGGATCGTCTTGACA 2453
Db 98 lysArgGlnArgSerIleSerAlaAspSerPheAspGlnArgAspProGlyThr----- 115
QY 2454 ATGAATAATGACGAATGAGCATGGAAGGCTGCAATCAGTTGAATCCGATTTTATCAAT 2513
Db 116 --ProAsnAspAspSerAspIleGlyGlnCysAsnSerAlaAspHisIleIlySerGln 134
QY 2514 GAATCTTTAAATAATCCTGCAATTTTCAGCATATTAAGTAAAGCGAGTAGGACCAATACCC 2573
Db 135 AspSerGlnHisThrProHisSerMetThr----- 144
QY 2574 GGAATCGAGTTGAGCGGGGACGGGAATTATTGACTGCCAACGCCAATGGAATCTCC 2633
Db 145 -----ProSerAsnAlaThrAlaProArgSerSerThrProSer 157
QY 2634 TCGGTTAGCAGTAATGTTGGATTACATGCAACAGCAAAATCAC-----ATA 2681
Db 158 HisGlyGlnThrThrAlaThrGluProThrProAlaGlnIlySerThrProAlaIlyValVal 177
QY 2682 TTGCTGTTTCAACTCAGCTGGCCAAAGGGCGGAGATCAGTTTAAAGCGTCAATT 2741
Db 178 TyrValPheSerThrGluMetAlaAsnIlyAlaIlaGluValLeuIlySgIlyGlnVal 197

QY 2742 CAAACTATATATGCGTATCACTGCACCTGCTACAAAAAGCTTCTGGAAGACTTT 2801
Db 198 GluThrIleValSerPheHis----- 204
QY 2802 TTTATGAAAAACCCCTTAAAGATTAAACAGTTACAGCGGCACAAATCCGTCGTATGCCA 2861
Db 205 ---IleGlnAsn---IleSerAsnAsnLysThrGluArgSerThrAla-----Pro 219
QY 2862 TGGATAGGCATGGCGGCAGGTTGGACTACTCTCTTAATCCTGTAGCCAAAATAACACAA 2921
Db 220 LeuAsnThrGlnIleSerAlaLeuArgAsnAspProLysProLeu-----Pro 235
QY 2922 CAGCAGCCACATACAAAGACCGTAGGCTTATTGAAACCCCAATTCAATCAACATGAAAAAC 2981
Db 236 GlnGlnProProAlaProAla-----AsnGlnAspGlnAsn 247
QY 2982 AGCAACGTAAGTACTGTAAGCGCGCTAGCAACTCTTTT----- 3020
Db 248 SerSerGlnAsnThrArgLeuGlnProThrProProIleProAlaProAlaProLysPro 267
QY 3021 -----GTCGACCAAGTCTGATCCTATGGCAACGAATACTGAATTGATG 3062
Db 268 AlaAlaProProArgProLeuAspArgLysSerPro---GlyValGluAsnLysLeuIle 286
QY 3063 TGCTGGGAAGCGGATCCTCAAAACACCACT-----AGGCTGCACAAAC 3107
Db 287 ProSerValGlySerProAlaSerSerThrProLeuProProAspGlyThrGlyProAsn 306
QY 3108 TCACGAATCATGTAGACAGTATCAGTACATCCAGCGAGTCACAGGCATTAAGATACTG 3167
Db 307 SerThrProAsnAsnArgAlaValThrProValSerGlnGlySerAsnSerSerAla 326
QY 3168 GAAGCAGCTGGCGTTGATTGGGACAGGTCAAAAGGAAGCGATCCTGGCCTGACAACT 3227
Db 327 AspProLysAlaProProProProProValSerSerGlyGluProProThrLeu---Gly 345
QY 3228 GAAAAACAACATTGTATCACTGCAAGAGTTAAGTTCAGACGAAAACTTACACCACAA 3287
Db 346 GluAsn-----ProAsp---GlyLeuSerGlnGlu 354
QY 3288 CAGCGCAACATCGGGAAGACAGTGGCAAAAAATAAAAAATGAATCAATTCTTTT 3347
Db 355 GlnLeuGlnHisArgGluArgSerLeuGlnThrLeuArgAspIleGlnArgMetLeuPhe 374
QY 3348 CCTGAAATGAGAATTCAAGTAGAGCTAATGTAAGCTCAGAGATACAAAAAATTCACAGA 3407
Db 375 ProAspGluLysGluPheThrGlyAlaGln----- 384
QY 3408 GATTTAATGATGGGATGTCGGGTGGCGGCGGATCTATTATAATCCGACGATGCCA 3467
Db 385 -----SerGlyGlyProGlnGlnAsnPro----- 392
QY 3468 CAACTGCATATGCGAGGTAACGCCAAATCGAGCTCTTATCGCGACAGTTACGACTT 3527
Db 392 ----- 392
QY 3528 TCGGAAGATGTAATGCATCCAGGGATGTTATATCAGATATGGTGCGTAATAGATGT 3587
Db 393 -----GlyValLeuAspGlyPro 398
QY 3588 AATATAATCAAAAAACAGTGTGCAATGTGATCTGAGTAGTGTTGTCACTGGAACA 3647
Db 399 GlnLysLysProGlnGlyProIleGlnAla----- 408
QY 3648 ACTGCAGCTGGAGTAATGTCAATATGCATTGCTCAAGCTCCGGCGCCCGAATGGCAAT 3707
Db 409 -----MetMetAlaGlnSerGlnSerLeuGlyLysGly--- 419
QY 3708 ATGATGGGAAGCTTAAGGATATAGCTAGCCTGTTGGCAACACAAAGCTGCAACGTATC 3767
Db 420 ---ProGlyProArgThrAspValGlyAlaProPheGlyProGlnGlyHisArgAspVal 438

QY 3768 GGAACGCCCCAGAT---ATGTTCTAAGGAAGTTTAAATCAAGATAGCCGAACC----- 3818
Db 439 ProPheSerProAspGluMetValProProSerMetAsnSerGlnSerGlyThrIleGly 458
QY 3819 -----CATTCACATCAAGGGGAGTTGCTCAATATGAGTGTGGAAGATTCACATCA 3872
Db 459 ProAspHisLeuAspHisMetThrProGlnGlnIleAlaThrLeuLysLeuGlnGlnGlu 478
QY 3873 TTTTTCGAAGAAGCCCTCAAGGGGGCAAGCCCAAGACAGTCACTGGAACCTGTAGTACCA 3932
Db 479 PheTyrgLysGlnLysArgArg-----LysGlnGlnGlnVal-----ValValGln 493
QY 3933 CAA-----CAGCAAAACCCCTTCTGATCTGTGGAAC 3965
Db 494 GlnCysSerLeuGlnAspMetMetValHisGlnHisGlyProArgGly----- 509
QY 3966 TCGTTAAACAACAGGTGCGACCCCTGCAAGGTCCACCTCTCTTACCAC----- 4016
Db 510 -----ValValArgGlyProProProProTyrgLysMetThrPro 522
QY 4017 -----TCCATCCAGAGATCTGCG 4034
Db 523 SerGlnGlyTyrAlaProGlyGlyThrGluProPheSerAspGlyIleAsnMetProHis 542
QY 4035 TCAGTACCAATAGCCACTCAATCGCCCAATCCCTCGAGTCCAAACAATCTA----- 4085
Db 543 SerLeuProProArgGlyMetAlaProHisProAsnMetProGlySerGlnMetArgLeu 562
QY 4086 -----TCTCTCCCGTCACCGCGG 4103
Db 563 ProGlyPheAlaGlyMetIleAsnSerGlnMetGlyProAsnValProAsnProAla 582
QY 4104 ACAACCGCAGCAGTCATGGGATG-----CCGACCACTCTCTAGCATG----- 4148
Db 583 SerArgProGlyLeuSerGlyValSerTrpProAspAspValProLysIleProAspGly 602
QY 4149 -----GATGGAACAGATCATTATCTGGA----- 4172
Db 603 ArgAsnPheProProGlyGlnGlyIlePheSerGlyProGlyArgGlyLysArgPhePro 622
QY 4173 -----TCTGTTCCGCAAGCTAATACTTCGACCGTTCAGCGCAGGC 4211
Db 623 AsnProGlnGlyLeuSerGlnGluMetPheGlnGlnGlnLeuAlaGluLysGlnLeuGly 642
QY 4212 ACAACA-----ACAGTGTCTTCAACAACAAGAACTGTTTCAGGCGACACCCCATCG 4265
Db 643 LeuProProGlyMetAlaMetGlnGlyIleArgProSerMetGluMetAsnArgMetIle 662
QY 4266 CCGTCAATCAAAATCGTAGTAAGATAACCGGATCGTCAAGCGTCTTACGCATAACTTA 4325
Db 663 ProGlySerGln---ArgHisMetGluProGlyAsnAsnProIlePheProArgIlePro 681
QY 4326 AGCAGCAACCCAGTACCCCTTATCTCATCTATCCCAAG----- 4367
Db 682 ValGlnGlyProLeuSerProSerArgGlyAspPheProLysGlyIleProProGlnMet 701
QY 4368 -----GAATTGAGTCTTTCCGTCAGTCCCTGCTGCTGATGTATATTGTTT 4415
Db 702 GlyProGlyArgGlnLeuGln---PheGlyMetValProSerGlyMet----- 716
QY 4416 AATTTTTTAAAGACAAATCAATATGAATTGCGTTAATAATAAGTTATATATTACATAA 4475
Db 717 -----LysGlyAspValAsnLeuAsn---ValAsn----- 725
QY 4476 CTCGAAATTTGATAGAAAAAATCAGAAATAGAAAAAATAATTTCGCGACCG--- 4532
Db 726 MetGlySerAsnSerGlnMetIleProGlnLysMetArgGluAlaGlyAlaGlyProGlu 745
QY 4532 ----- 4532
Db 746 GluMetLeuLysLeuArgProGlyGlySerAspMetLeuProAlaGlnGlnLysMetVal 765
QY 4533 -----CCCATCATTTCTTG 4547

Db 766 ProLeuProPheGlyGluHisProGlnGlnGluTyrglyMetGlyProArgProPheLeu ||| ||||| 785
QY 4548 -----AATCCAATTCTGGAGTGATGTTAGATAA----- 4579
Db 786 ProMetSerGlnGlyProGlySerAsnSerGlyLeuArgAsnLeuArgGluProIleGly 805
QY 4580 -----TCTACTATTAAATTAAACACGAAATTCATATCCGTTAATTGAAAT 4627
Db 806 ProAspGlnArgThrAsnSerArgLeuSerHisMetProProLeuProLeu----- 822
QY 4628 CACTATTGTTTAATAGAATTAATAATATGTTATTATAATTCTACAGGTGATAAC 4687
Db 823 -----AsnProSerSerAsn 827
QY 4688 ATGAAAGTAGCGACCAAGCCACAGGTCAGCGGTCAACAGTAATAAGTCTAATAGAG 4747
Db 828 ProThrSerLeuAsnThrAlaProProValGlnArg-----GlyLeuGlyArg 843
QY 4748 GCAAAATAAGATGTACGATTGCTGCATCC-----AGTCCGTGTTTAAACCCGCAT 4798
Db 844 LysProLeuAspIleSerValAlaGlySerGlnValHisSerProGlyIleAsnPro--- 862
QY 4799 CCACATATGCAAAAGCAATTCAATTTCAGCATTAACGCTATAAAATGGGCTTCACCAAT 4858
Db 863 -----LeuLysSerProThrMetHisGlnValGlnSerProMetLeuGlySerProSer 880
QY 4859 ATACAGATGAGGTAAATATTAAATATTTATTAAAGTTTGTGTTAATTATCTTC 4918
Db 881 GlyAsnLeu----- 883
QY 4919 TTTTTCAGCGTCAGCATCAGCCAGAGTGATCCGTAATTAGTCGC----- 4969
Db 884 -----LysSerProGlnThrProSerGlnLeuAlaGlyMetLeuAla 897
QY 4970 -----GCTCCGATAATATTCGCTAAATCCCAATAGTGGCAATCGGCCGC----- 5015
Db 898 GlyProAlaAlaAlaAlaSerIleLysSerProProValLeuGly-SerAlaAlaAlaSe 917
QY 5016 -----ACCAACAAGA 5026
Db 917 rProValHisLeuLysSerProSerLeuProAlaProSerProGlyTrpThrSerSerPr 937
QY 5027 TGACCCAAACTTCGATCCCAATCTTCTTTGGCACAATGTCCCAACAATAAGATT 5086
Db 937 cLysProProLeuGlnSer-----ProGlyIleProProAsnHisIleVal 952
QY 5087 GCGTGTCCAGCATGGGTAGTCCAGCCGGAAGTGGTGTATGACGATGAGGGGGTCCG- 5145
Db 952 aProLeu-ThrMetAlaSerProAlaMetLeuGlyAsnValGlnSerGlyGlyProPro 972
QY 5146 -----GGACCGTCCGACATCAATATTGAG----- 5169
Db 972 roProThrAlaSerGlnProAlaSerValAsnIleProGlySerLeuProSerSerThrP 992
QY 5169 ----- 5169
Db 992 roTyTrpThrMetProProGluProThrLeuSerGlnAsnProLeuSerIleMetMetSera 1012
QY 5170 -----CATGAATAATTTCCGGAC 5188
Db 1012 rgMetSerLysPheAlaMetProSerSerThrProLeuTyTrpHisAspAlaIleLysThrV 1032
QY 5189 TAGATGATCAGGAATAGATACC-----ATAAATCAAAATAACTGTCAATCAATGATG 5242
Db 1032 aAlaSerSerAspAspAspSerProProAlaArgSerProAsnLeuProSerMetAsn- 1051
QY 5243 TCGTAATGAAGTCAATGGTCCCGAATGCTGAATCCTTAAATGTGCGTAGCAGCGCTC 5302
Db 1052 --AsnMetProGlyMetGlyIleAsnThrGlnAsnProArgIle-----SerGlyP 1068
QY 5303 CAAT----- 5307

Db 1068 roAsnProValProMetProThrLeuSerProMetGlyMetThrGlnProLeuSerH 1088
QY 5308 -----GACCGCCTGGCTTAAATCCTAATTCGCCCAATGGTGATTA 5350
Db 1088 IsSerAsnGlnMetProSerProAsnAlaValGlyProAsnIleProProHisGlyVal- 1107
QY 5351 GAGAGAATTCATAGGCTGGCTGGCTGGCTCAGCAACTCTTCAAACTTTCAGGGGTTG 5410
Db 1108 -----PrometGlyProGlyLeuMetSerHisAsnProIleMetGlyHisGlySerG 1125
QY 5411 TTCCACCTGGTCCAGATGATGGGTGCAATGCCAGTCAATTTGTTGCAATTTCATC 5470
Db 1125 InGluProPrometValProGlnGlyArgMet-----GlyPheProGlnGlyPheProP 1143
QY 5471 CGAATATTGAGTAAAGCGGAGTACCCCAACCAATCAATACATGATCCAGTAAGGCAC 5530
Db 1143 ro-----ValGlnSerProProGlnGlnValP 1152
QY 5531 AGAAGCCCAACAATAACAATGAGAGCTAAATGTCGAATGCCACCTAGTCTCG 5590
Db 1152 roPheProHisAsnGlyProSerGlyGlyGlnGlySer-----PheProGlyGlyMetG 1170
QY 5591 AATTTTGCAGAGGTACGTAACCTCAATGGGTGCTAGGCAATGGGTGCGCAATAT 5650
Db 1170 Lyphe-----ProGlyGlyGlyProLeuGlyArgProSerAsnLeu- 1183
QY 5651 GCCCACCATCAGCCAGCAGCGGTACTCTGGAATGCCAGATTGATGGCGGACCAAGAG 5710
Db 1184 --ProGlnSerSerAlaAspAlaAlaLeuCysLysProGlyGlyProGlyGlyProAspS 1203
QY 5711 CCGAGGTATGCTAATGAATCT----- 5733
Db 1203 erPheThrValLeuGlyAsnSerMetProSerValPheThrAspProAspLeuGlnGluV 1223
QY 5733 ----- 5733
Db 1223 alIleArgProGlyAlaThrGlyIleProGluPheAspLeuSerArgIleIleProSerG 1243
QY 5734 -----TCCGAGAGACAACACCAGAACAGATCA 5761
Db 1243 IuLysProSerGlnThrLeuGlnTyTrpPheProArgGlyGluValProGlyArgLysGlnP 1263
QY 5762 CAACAATCCTGGGCGCAAGCAATGATTAATCTTTTCAGAATTGCAATCAATGTCTA 5821
Db 1263 roGlnGlyProGlyPro-----GlyPheSerHisMetGln-----GlyM 1276
QY 5822 TTGTTGACGAAGAGGT-----GATTAACCGCGCATGACGATCAATGA 5866
Db 1276 etMetGlyGlnGlnAlaProArgMetGlyLeuAlaLeuProGlyMetGlyGlyProGlyP 1296
QY 5867 ATATTGTCACACCATGTATGATAAGGCGCATGCTCA-----CATGCCA 5911
Db 1296 roValGlyThrProAspIleProLeuGlyThrAlaProSerMetProGlyHisAsnProM 1316
QY 5912 TGCGGCCCAAT-----GTATGGGTGCGCGGATGCCACCCGTTAACAGGCAATTCAGT 5965
Db 1316 etArgProProAlaPheLeuGlnGlnGlyMetMetGlyProHisHisArgMetMetSerP 1336
QY 5966 TTGCACAGTCATCGGATGATATTGACTGTGTGGGATCCGTCATCATTTTTCATAACG 6025
Db 1336 roAlaGlnSerThr-----MetProGlyGlnProThrLeuMetSerAsnProA 1352
QY 6026 CTTCTGCAACAGCGCTGGACACACATGTTTGGATTCAGCACACAGGCGCAATCAGCCTA 6085
Db 1352 IaAlaAla----- 1354
QY 6086 AGACACAACACATAAAGAACATACCTAGTGAATGTGTCAAAACCAATCGGACTTCAG 6145
Db 1355 -----ValGlyMetIlePro--GlyLysAspArgGlyProAlaGlyLeuTyTr 1370
QY 6146 TGGACAAGGCGAGATCCAATGTCATGGGCAAGACATGCGAGGTCAGTCTTAAATTG 6205
Db 1370 hrHisProGlyProValGlySerProGlyMetMetMetSerMet--GlnGlyMetMetG 1389

```

QY      6206 GACCTACTAATAATAATTA-----ATGTCACCTGCCGGAAGTGTCACTGCTA 6253
          |||||      |||      ::      |||      :::::||||
Db      1389 lYPRoGInGlnAsnIlEmEtIlEProProGlnMetArgProArgGly--MetaIaIaA 1408
          ||:::      |||      ::      |||||      ::
QY      6254 CTAACGGTGTCTCTGGCATCAATTTCGTAGTCCCTCTTCT 6294
          ||:::      |||      ::      |||||      ::
Db      1408 spvAlGlyMetGlyGlyPheSer---GlnGlyProGlyAsn 1420

RESULT 10
ID      ABW01534 standard; protein; 1426 AA.
XX
XX      AC      ABW01534;
XX      DT      15-JAN-2004 (first entry)
XX      DE      Human Igs/bcl9 protein.
XX
XX      KW      Legless protein; Igs; cell fate disorder; blood disease; gene therapy;
XX      KW      cancer; tissue regeneration; tissue repair; cytostatic.
XX      OS      Homo sapiens.
XX      PN      US2003114413-A1.
XX      PD      19-JUN-2003.
XX      PF      19-DEC-2002; 2002US-00322579.
XX      PR      28-JUL-2000; 2000US-0221502P.
XX      PR      27-JUL-2001; 2001US-00915543.
XX      PA      (UYZU-) UNIV ZURICH.
XX
XX      PI      Basler K, Brunner E, Froesch B, Kramps T, Peter O;
XX      DR      WPI; 2003-829432/77.
XX      DR      N-PSDB; AAD62642.
XX
XX      PT      Novel Igs polypeptide useful for isolation of Igs-binding proteins,
XX      PT      diagnosing disorders of cell fate, treating diseases such as cancer.
XX
XX      PS      Example 2; Fig 8B; Opp; English.
XX
CC      The invention relates to novel legless (Igs) proteins and polynucleotides
CC      encoding such proteins. Igs sequences are useful for the treatment of
CC      disorders of cell fate such as differentiation or proliferation. The
CC      invention is used to treat blood disease or a cancerous condition
CC      characterised by over-stimulation of the Wnt pathway such as colon,
CC      breast, head and neck, brain, thyroid, medulloblastoma or skin cancer and
CC      is administered to prevent progression from a pre-neoplastic or non-
CC      malignant condition to a neoplastic or malignant state. It is
CC      administered to promote tissue regeneration and repair. The invention is
CC      also useful in the therapy of diseases cost by an over-activation of Wg
CC      pathway. It is useful for reducing Igs gene expression in an invertebrate
CC      or vertebrate organism or an invertebrate or vertebrate cell line. The
CC      invention is also useful in gene therapy. The present sequence is human
CC      Igs/bcl9 protein used in the invention
XX
XX      SQ      Sequence 1426 AA;

Alignment Scores:
Pred. No.:      4.69e-17      Length:      1426
Score:      312.00      Matches:      313
Percent Similarity:      30.34%      Conservative:      177
Best Local Similarity:      19.38%      Mismatches:      519
Query Match:      2.59%      Indels:      607
DB:      7      Gaps:      73

US-10-664-859-1 (1-6909) x ABW01534 (1-1426)
QY      2397 AAAATTGAAAGA--ATTTCAAACGACAGTACCACGGAAGAAAAAGGATCGTCTTGACA 2453

```

| | | | |
|----|------|--|------|
| Db | 98 | LysArgGluArgSerLeuSerAlaAspSerPheAspGlnArgAspProGlyThr----- | 115 |
| QY | 2454 | ATGAATAATGACGAATGAGCATGGAAGGCTGCATCAGTTGAATCCGATTTATCAAT | 2513 |
| Db | 116 | ---ProAsnAspAspSerAspIleLeuGluCysAsnSerAlaAspHisIleLeuSerGln | 134 |
| QY | 2514 | GAATCTTAAATAATCTCTCCAATTTTCGAGCATATTAGTAAGCGAGTAGGACCAATACC | 2573 |
| Db | 135 | AspSerGlnHisThrProHisSerMetThr----- | 144 |
| QY | 2574 | GGAATCGAGTTGGACGGGGGACGGAAATTATTGACTGCCAACGCCAATGGAATCTCC | 2633 |
| Db | 145 | -----ProSerAsnAlaThrAlaProArgSerSerThrProSer | 157 |
| QY | 2634 | TCGGGTAGCAGTAATGTTTGGATTACATGCAACAGCAAAATCAC-----ATA | 2681 |
| Db | 158 | HisGlyGlnThrThrAlaThrGluProThrProAlaGlnIleThrProAlaIleValVal | 177 |
| QY | 2682 | TTCCGTGTTTTCACACTGACCTGGCCAAACAAGGGCCGAATCAGTTTAAAGCGTCAATT | 2741 |
| Db | 178 | TyrValPheSerThrGluMetAlaAsnLysAlaIleAlaValLeuLysGlyGlnVal | 197 |
| QY | 2742 | CAAACTATTATTGCCGTATCACTGCACACTCAGCCTGCTCAAAAAGCTTCCTGGAAGACTTT | 2801 |
| Db | 198 | GluThrIleValSerPheHis----- | 204 |
| QY | 2802 | TTTATGAAAAACCCCTTTAAAGATTTAACAAGTTACAGCGGCACAATTCCTCGGTATGCCA | 2861 |
| Db | 205 | ---IleGlnAsn--IleSerAsnAsnLysThrGluArgSerThrAla-----Pro | 219 |
| QY | 2862 | TGGATAGGCATGGGGCAGGTGGACTAACTCTCTTAATCTCTGTAGCCAAATTAACAA | 2921 |
| Db | 220 | LeuAsnThrGlnIleSerAlaLeuArgAsnAspProLysProLeu-----Pro | 235 |
| QY | 2922 | CAGCAGCCACATACAAGAACCGTAGCCTATTGAAACCCCAATTCAATCAACATGAAAAAC | 2981 |
| Db | 236 | GlnGlnProProAlaProAla-----AsnGlnAspGlnAsn | 247 |
| QY | 2982 | AGCAACGTAAGTCTGTAAAGCGCGCTAGCAACTCTTT----- | 3020 |
| Db | 248 | SerSerGlnAsnThrArgLeuGlnProThrProProIleProAlaProAlaProLysPro | 267 |
| QY | 3021 | -----GTCGACCAGTCTGATCCTATGGGCAACGAAACTGAATTGATG | 3062 |
| Db | 268 | AlaAlaProProArgProLeuAspArgGluSerPro--GlyValGluAsnLysLeuIle | 286 |
| QY | 3063 | TGCTGGGAAGCGGATCCTCAAAACACCAGT-----AGTCTGACAAAAAC | 3107 |
| Db | 287 | ProSerValGlySerProAlaSerSerThrProLeuProProAlaGlyThrGlyProAsn | 306 |
| QY | 3108 | TCACGAAATCATGTAGACAGTATCATACATCCAGCAGTCCACAGCGCAATAAAGTAACTG | 3167 |
| Db | 307 | SerThrProAsnAsnArgAlaValThrProValSerGlnGlySerAsnSerSerSerAla | 326 |
| QY | 3168 | GAAGCAGCTGCGTTGATTGGGACAGTCAACAAGAAAGCGATCCTGCGCTGCAACT | 3227 |
| Db | 327 | AspProLysAlaProProProProProValSerSerGlyGluProProThrLeu--Gly | 345 |
| QY | 3228 | GAAAACAACATTGTATCTACTGCAAGAGTTAAGTTCACAGACGAAAAACCTTACACCAAA | 3287 |
| Db | 346 | GluAsn-----ProAsp--GlyLeuSerGlnGlu | 354 |
| QY | 3288 | CAGCGGCAACATCGGGAAGAACAAGTGGGCAAAAAATAAAAAATGAATCAATTTCTTTT | 3347 |
| Db | 355 | GlnLeuGluHisArgGluArgSerLeuGlnThrLeuArgAspIleGlnArgMetLeuPhe | 374 |
| QY | 3348 | CCTGAAATGAGAAATTCAGTAGGAGCTAATGTAAAGTCCACAGATAACAAAAATTCACAGGA | 3407 |
| Db | 375 | ProAspGluLysGluPheThrGlyAlaGln----- | 384 |
| QY | 3408 | GATTTAATGATGGGATGTCGGGTGGCGAGCGGATCTATTATAATCCGACGATGCGA | 3467 |

Db 385 -----SerglycylProglInGlnAsnPro----- 392
QY 3468 CAACTGCATATGCCAGGTAAACGCCAAATCGAGCTTATCGCGACAAAGTTCAAGACTT 3527
Db 392 ----- 392
QY 3528 TCGAAGATGTAATGCATCCAGGGAGTGTATATCAGATATGGTCCGTAATAGATGT 3587
Db 393 -----GlyValLeuAspGlyPro 398
QY 3588 AATAATAATCAAAAAACAGTGTGCAATGTGATCTGAGTAGGTGTGTCTACTGGAACA 3647
Db 399 GlnLysLysProGlnGlyProIleGlnAla----- 408
QY 3648 ACTGCAGCTGGAGTAATGTCAATATGTCATTGCTCAAGCTCCGGCGCCCGAATGGCAAT 3707
Db 409 -----MetMetAlaGlnSerGlnSerLeuGlyLysGly--- 419
QY 3708 ATGATGGAAAGCTCTACGATATGCTAGCCTGTTTGGCAACACAAAGCTGCAACGTCATC 3767
Db 420 ---ProGlyProArgThrAspValGlyAlaProPheGlyProGlnGlyHisArgAspVal 438
QY 3768 GGAACGCCCCAGAT---ATGTCTAAGAAAGTTTAAATCAAGATAGCCGAAC----- 3818
Db 439 ProPheSerProAspLumetValProProSerMetAsnSerGlnSerGlyThrIleGly 458
QY 3819 -----CATTCACATCAAGGGGGAGTTGCTCAATGAGTGTGCAAGATTCACATCAA 3872
Db 459 ProAspHisLeuAspHisMetThrProGlnGlnIleAlaTrpLeuLysLeuGlnGlnGlu 478
QY 3873 TTTTTCGAAGAACCCCTCAAGGGGGGCAAGCCCAAGTCAAGTCACTGGAAGTGTAGTACCA 3932
Db 479 PheTyrgLysGlnLysArgArg-----LysGlnGlnGlnVal-----ValValGln 493
QY 3933 CAA-----CAGCAAAACCCTTCTGGATCTGTGGAAC 3965
Db 494 GlnCysSerLeuGlnAspMetMetValHisGlnHisGlyProArgGly----- 509
QY 3966 TCGTTAAACAACAGGTGCGACCCCTGCAAGGTCCACCTCTCTTACCAAC----- 4016
Db 510 -----ValValArgGlyProProProProTyrgLumetThrPro 522
QY 4017 -----TCCATCCAGAGATCTGCG 4034
Db 523 SerGlnGlyTrpAlaProGlyGlyThrGluProPheSerAspGlyIleAsnMetProHis 542
QY 4035 TCAGTACCAATAGCCACTCAATCGCCCAATCCCTCGAGTCCAAACAATCTA----- 4085
Db 543 SerLeuProArgGlyMetAlaProHisProAsnMetProGlySerGlnMetArgLeu 562
QY 4086 -----TCTCTCCCGTCACCGCGG 4103
Db 563 ProGlyPheAlaGlyMetIleAsnSerGlnLumetGlnGlyProAsnValProAsnProAla 582
QY 4104 ACAACCGCAGACATCAGGATTG-----CCGACCAACTCTCTAGCATG----- 4148
Db 583 SerArgProGlyLeuSerGlyValSerTrpProAspAspValProLysIleProAspGly 602
QY 4149 -----GATGGAACAGGATCATTAATCTGA----- 4172
Db 603 ArgAsnPheProProGlyGlnGlyIlePheSerGlyProGlyArgGlyLysArgPhePro 622
QY 4173 -----TCTGTTCCGCAAGCTAATACTTCGACCGTTCAAGGACGAGC 4211
Db 623 AsnProGlnGlyLeuSerGlnGlnLumetPheGlnGlnGlnLeuAlaGlnLysGlnLeuGly 642
QY 4212 ACAACA-----ACAGTGTCTTCAGCAACAAGAACTGTTTTCAGGACAGACACCCCATCG 4265
Db 643 LeuProProGlyMetAlaMetGlnGlyIleArgProSerMetGlnMetAsnArgMetIle 662
QY 4266 CCGTCAAAATCAAAATCGTAGTAAGAAATACCGGATCGTCAAGCGTTCCTTACGATAAATTA 4325
Db 663 ProGlySerGln---ArgHisMetGlnProGlyAsnAsnProIlePheProArgIlePro 681

QY 4326 AGCAGCAACCAAGTACCCCTTATCTCATCTATCCCAAG----- 4367
Db 682 ValGlnGlyProLeuSerProSerArgGlyAspPheProLysGlyIleProProGlnMet 701
QY 4368 -----GAATTGAGTCTTTGGCTCAGTCCCTCTGCTGGTATGTATATTGTTT 4415
Db 702 GlyProGlyArgGlnLeuGlu---PheGlyMetValProSerGlyMet----- 716
QY 4416 AATTTTTTAAAGACAAATCAAAATGAAATGTCGTTAATTAATTAATTAATTAATTAATTA 4475
Db 717 -----LysGlyAspValAsnLeuAsn---ValAsn----- 725
QY 4476 CTCGAAATTTGATAGAAAAAATCAGGAATAGAAAAAATTAATTAATTTCCGACCG--- 4532
Db 726 MetGlySerAsnSerGlnMetIleProGlnLysMetArgGlnAlaGlyAlaGlyProGlu 745
QY 4532 ----- 4532
Db 746 GlnMetLeuLysLeuArgProGlyGlySerAspMetLeuProAlaGlnLysMetVal 765
QY 4533 -----CCCATCCATTCTTG 4547
Db 766 ProLeuProPheGlyGlnHisProGlnGlnGlnTyrgLysMetGlyProArgProPheLeu 785
QY 4548 -----AATCCAAATTCTCGAGTGTGTTAGATAA----- 4579
Db 786 PrometSerGlnGlyProGlySerAsnSerGlyLeuArgAsnLeuArgGluProIleGly 805
QY 4580 -----TCTACTATTAAATTAACACGAAAAATTCATATCCGTTAATTGAAAT 4627
Db 806 ProAspGlnArgThrAsnSerArgLeuSerHisMetProProLeuProLeu----- 822
QY 4628 CACTATGTGTTAATAAGAAATTAATAATGTTTATTATTAATATTCTTCAAGGTGATAAC 4687
Db 823 -----AsnProSerSerAsn 827
QY 4688 ATGAAAAGTAGCCGCAACCAAGCCACAGGTCACGGTCAACAGTAATAGTCTAATAGAG 4747
Db 828 ProThrSerLeuAsnThrAlaProProValGlnArg-----GlyLeuGlyArg 843
QY 4748 GCAATAAAGATGTACGATTGCTGCATCC-----AGTCTGCTTTTAAACCCGCAT 4798
Db 844 LysProLeuAspLysSerValAlaGlySerGlnValHisSerProGlyIleAsnPro--- 862
QY 4799 CCACATATGCAAGCAATTCAAATTCAGCATTAAGCCCTATAAAATGGGCTTACCAAT 4858
Db 863 -----LeuLysSerProThrMetHisGlnValGlnSerPrometLeuGlySerProSer 880
QY 4859 ATACAGATGAGGTAATATATTTAATATTATTAAAGTTTGTGTTAATTATCTTC 4918
Db 881 GlyAsnLeu----- 883
QY 4919 TTTTTCAGCGTCAAGCATCAGCGCAAGGTGATCCGTACAATTAGTCGC----- 4969
Db 884 -----LysSerProGlnThrProSerGlnLeuAlaGlyMetLeuAla 897
QY 4970 -----GCTCCGATAATATTCGCTAAATCCCAATAGTGGCAATCGCGCC----- 5015
Db 898 GlyProAlaAlaAlaAlaSerIleLysSerProProValLeuGly-SerAlaAlaAlaLase 917
QY 5016 -----ACCAACAAGA 5026
Db 917 rProValHisLeuLysSerProSerLeuProAlaProSerProGlyTrpThrSerSerPr 937
QY 5027 TGACCCAAACTTCGATCCCAATCTCTTTTGGACAAATGTCCCAACAACAACTAACAAGTT 5086
Db 937 LysProProLeuGlnSer-----ProGlyIleProProAsnHisLysAl 952
QY 5087 GCGTGTCCAGATGGGTAGTCCAGCCGGAACCTGGTGTATGACGATGAGGGGGTCCG- 5145
Db 952 aProLeu-ThrMetAlaSerProAlaMetLeuGlyAsnValGlnSerGlyGlyProProp 972


```

QY 5146 -----GGACCGTCCGACATCATATTGAG----- 5169
Db 972 koProThrAlaSerGlnProAlaSerValAsnIleProGlySerIleuProSerSerThrP 992
QY 5169 ----- 5169
Db 992 koTyThrMetProProGluProThrIleuSerGlnAsnProIleuSerIleMetMetSerA 1012
QY 5170 -----CATGGAATAATTTCGGGAC 5188
Db 1012 rgMetSerIlySPheAlaMetProSerSerThrProLeuTyThrIleAspAlaIleLysThrV 1032
QY 5189 TAGATGATCAGAGATAGATACC-----ATAAATCAAAATTAAGTGTCAATTCATGAATG 5242
Db 1032 alAlaSerSerAspAspAspSerProProAlaArgSerProAsnIleuProSerMetAsn- 1051
QY 5243 TCGTAATGAATCAATGGGTCCCGGATGCTGAATCCTAAATGTGCGTAGCAGCGGTC 5302
Db 1052 --AsnMetProGlyMetGlyIleAsnThrGlnAsnProArgIle-----SerGlyP 1068
QY 5303 CAAAT----- 5307
Db 1068 roAsnProValValProMetProThrIleuSerProMetGlyMetThrGlnProIleuSerH 1088
QY 5308 -----GGACCGCTGGCTTTAATCCTAATTCGCCCAATGGTGATTA 5350
Db 1088 isSerAsnGlnMetProSerProAsnAlaValGlyProAsnIleProProHISGlyVal- 1107
QY 5351 GAGAGAATTCATAGGGTGTGGTGGCTGAGCAAACTCTTCAACTTTCAAGGGTTG 5410
Db 1108 -----PrometGlyProGlyLeuMetSerHISAsnProIleMetGlyHISGlySerG 1125
QY 5411 TTCCACCTGGTGCCAGATGATGGTGGATGCCAGTCAATTTGGTTCGAATTTCAATC 5470
Db 1125 lngluProPrometValProGlnGlyArgMet-----GlyPheProGlnGlyPheProP 1143
QY 5471 CGAATATTCAAGTAAAGCGAGTACCCCAAACACCATACATACATGCCAGTAAGGGCAC 5530
Db 1143 ro-----ValGlnSerProProGlnGlnValP 1152
QY 5531 AGAAGCCCAACAACATATACACAATGGAGCTAATATGTGCGAATGCCACTGATCTGG 5590
Db 1152 roPheProHISAsnGlyProSerGlyGlyGlnGlySer-----PheProGlyGlyMetG 1170
QY 5591 AATTTTGCAGAGGTACGCTAACCTCAATGGGTGTAGGCAATGGGTGCCCAATAT 5650
Db 1170 lyPhe-----ProGlyGlnGlyProLeuGlyArgProSerAsnIleu- 1183
QY 5651 GCCCACCATCAGCCAGCGAGCTACTCTCTGAATGCCAGATTGATGCCGCGGACCGAGAG 5710
Db 1184 --ProGlnSerSerAlaAspAlaAlaIleuCylysProGlyGlyProGlyGlyProAspS 1203
QY 5711 CCGAGGTATGCTAATGAATCT----- 5733
Db 1203 erPheThrValIleuGlyAsnSerMetProSerValPheThrAspProAspIleuGlnIuV 1223
QY 5733 ----- 5733
Db 1223 alIleArgProGlyAlaThrGlyIleProGluPheAspIleuSerArgIleIleProSerG 1243
QY 5734 -----TCCGAGAGCAACACAGACAAGATCA 5761
Db 1243 luLysProSerGlnThrLeuGlnTyThrPheProArgGlyGluValProGlyArgIlySglnP 1263
QY 5762 CAAACAATCTCTGGGCAAGCAATGTATTACTTTTCAGAAATTCGAATCAATGTCTA 5821
Db 1263 roGlnGlyProGlyPro-----GlyPheSerHISMetGln-----GlyM 1276
QY 5822 TTGTTGACGAGAGGGT-----GATTACCCGCGCATGACGATCAATGA 5866
Db 1276 etMetGlyGlnGlnAlaProArgMetGlyLeuAlaIleuProGlyMetGlyGlyProGlyP 1296
QY 5867 ATATTGGTCAACCATCTATGATAAGGGGCGATGCGTCA-----CATGCCA 5911

```

```

Db 1296 roValGlyThrProAspIleProIleuGlyThrAlaProSerMetProGlyHISAsnProM 1316
QY 5912 TGGGCGCAAAT-----GTAATGGGTGCGCGGATGCCACCGGTAAACAGCAAAATTCAGT 5965
Db 1316 etArgProProAlaPheIleuGlnGlnGlyMetMetGlyProHISISArgMetMetSerP 1336
QY 5966 TTGCACAGTCATCGGATGGTATGTAAGTGTGTGCGGGATCCGTCATCATTTTTCACSTAAG 6025
Db 1336 roAlaGlnSerThr-----MetProGlyGlnProThrIleuMetSerAsnProA 1352
QY 6026 CTTCCTGCCAACAGCGCTGGACCAACATGTTTGATCAGCAACAAGCCCAATCAGCCTA 6085
Db 1352 laAlaAla----- 1354
QY 6086 AGACACAACACATTAAGAACATACCTAGTGAATGTGTCAAAACCAATCGGACTTGCAG 6145
Db 1355 -----ValGlyMetIlePro--GlyLysAspArgGlyProAlaGlyLeuTyrt 1370
QY 6146 TGGCACAAGGCGAGATCCAACTGCATGGGCAAGACATGCCAGGCTCAGTCTTTAATTG 6205
Db 1370 htrHISProGlyProValGlySerProGlyMetMetMetSerMet--GlnGlyMetMetG 1389
QY 6206 GACCTACTAATAATAATTTA-----ATGTCAACTGCCGGAAGTGCAGTGTCTA 6253
Db 1389 lyProGlnGlnAsnIleMetIleProProGlnMetArgProArgGly--MetAlaAlaA 1408
QY 6254 CTACCGTGTCTCTGGCATCAATTCGTAGTCCCTCTTCT 6294
Db 1408 spValGlyMetGlyGlyPheSer--GlnGlyProGlyAsn 1420

RESULT 11
ADJ70152
ID ADJ70152 standard; protein; 1426 AA.
XX
AC ADJ70152;
XX
DT 06-MAY-2004 (first entry)
XX
DE Human heat mitochondrial protein as a therapeutic target SeqID1958.
XX
KW mitochondrial; human; screening assay; diabetes mellitus;
KW Huntington's disease; osteoarthritis;
KW Leber's hereditary optic neuropathy; LHON;
KW mitochondrial encephalopathy lactic acidosis and stroke; MELAS;
KW myoclonic epilepsy ragged red fibre syndrome; MERRF; cancer;
KW neuroprotective; nootropic; antidiabetic; anticonvulsant; antiarthritic;
KW osteopathic; ophthalmological; cyostatic.
XX
OS Homo sapiens.
XX
PN WO2003087768-A2.
XX
PD 23-OCT-2003.
XX
PF 04-APR-2003; 2003WO-US010870.
XX
PR 12-APR-2002; 2002US-0372843P.
PR 17-JUN-2002; 2002US-0389987P.
PR 20-SEP-2002; 2002US-0412418P.
XX
PA (MITO-) MITOKOR.
PA (BUCK-) BUCK INST AGE RES.
XX
PI Ghosh SS, Fahy ED, Zhang B, Gibson BW, Taylor SW, Glenn GM;
PI Warnock DE;
XX
DR WPI; 2003-845369/78.
XX
PT Identifying a mitochondrial target for drug screening assays and for
PT treating diseases associated with altered mitochondrial function,
PT comprises detecting a modified polypeptide in a sample and correlating
PT with the disease.

```


QY 4017 -----TCCATCCAGAGATCTGCG 4034
Db 523 SerGluGlyTrpAlaProGlyGlyThrGluProPheSerAspGlyIleAsnMetProHis 542
QY 4035 TCAGTACCAATAGCCACTCAATCGCCCAATCCCTCGAGTCCAAACAATCTA----- 4085
Db 543 SerLeuProProArgGlyMetAlaProHisProAsnMetProGlySerGlnMetArgLeu 562
QY 4086 -----TCTCTCCCGTCACCGCGG 4103
Db 563 ProGlyPheAlaGlyMetIleAsnSerGluMetGluGlyProAsnValProAsnProAla 582
QY 4104 ACAACCGCAGCAGTCATGGGATTG-----CCGACCAACTCTCTAGCATG----- 4148
Db 583 SerArgProGlyLeuSerGlyValSerTrpProAspAspValProIlySileProAspGly 602
QY 4149 -----GATGAACAGAGATCATTTATCTGA----- 4172
Db 603 ArgAsnPheProProGlyGlnGlyIlePheSerGlyProGlyArgGlyGluArgPhePro 622
QY 4173 -----TCTGTCCGCAAGCTAATACTTCGACGGTTCAGGACGC 4211
Db 623 AsnProGlnGlyLeuSerGluGluMetPheGlnGlnGlnLeuAlaGluIlySglnLeuGly 642
QY 4212 ACAACA-----ACAGTCTCTCAGCAACAAGAACTGTTTTCAGGACAGACACCCCATCG 4265
Db 643 LeuProProGlyMetAlaMetGluGlyIleArgProSerMetGluMetAsnArgMetIle 662
QY 4266 CCGTCAATCAAAATCGTAGAGAATACCGGATCGTCAAGCGTTCTTACGCATAACTTA 4325
Db 663 ProGlySerGln--ArgHisMetGluProGlyAsnAsnProIlePheProArgIlePro 681
QY 4326 AGCAGCAACCAAGTACCCCTTATCTCATCTATCCCAAG----- 4367
Db 682 ValGluGlyProLeuSerProSerArgGlyAspPheProIlySgIlyIleProProGlnMet 701
QY 4368 -----GAATTGAGTCTTTCGGTCAGTCTCTGCTGCTGTAATGTTATTTGTTT 4415
Db 702 GlyProGlyArgGluLeuGlu--PheGlyMetValProSerGlyMet----- 716
QY 4416 AATTTTAAAGACAATCAATATGAATGCGTTAATATAATAGTTATATATACATAA 4475
Db 717 -----LysGlyAspValAsnLeuAsn--ValAsn----- 725
QY 4476 CTCGAAATTTGATAGAAAAATCAGGAATGAAAAAATAATTATTTCCGACCG--- 4532
Db 726 MetGlySerAsnSerGlnMetIleProGlnIlyMetArgGluAlaGlyAlaGlyProGlu 745
QY 4532 ----- 4532
Db 746 GluMetLeuIlyLeuArgProGlyGlySerAspMetLeuProAlaGlnGlnIlyMetVal 765
QY 4533 -----CCCATCCATTCTTG 4547
Db 766 ProLeuProPheGlyGluHisProGlnGlnGluTyrglyMetGlyProArgProPheLeu 785
QY 4548 -----AATCCAATTTCTGAGTGAATTGTAGAGATAA----- 4579
Db 786 PrometSerGlnGlyProGlySerAsnSerGlyLeuArgAsnLeuArgGluProIleGly 805
QY 4580 -----TCTACTATTAAATTAACGAGAAATTCATATCCGTTAATTGAAAT 4627
Db 806 ProAspGlnArgThrAsnSerArgLeuSerHisMetProProLeuProLeu----- 822
QY 4628 CACTATTGTTTAATAGAATAAATAATATGTTTATTATATATTCTACAGGTGATAAC 4687
Db 823 -----AsnProSerSerAsn 827
QY 4688 ATGAAAGTAGCGCAGCAAGCCACAGGTCAGCGGTCAACAGTAATAAGTCTAATAGAG 4747
Db 828 ProThrSerLeuAsnThrAlaProProValGlnArg-----GlyLeuGlyArg 843

QY 4748 GCAATAAAGATGTACGATTGCTGCATCC-----AGTCTGCTTTTAACCCGCAT 4798
Db 844 LysProLeuAspIleSerValAlaGlySerGlnValHisSerProGlyIleAsnPro--- 862
QY 4799 CCACATATGCAAGCAATTCAAATTACGATTAAACGCTTAAATGAGCTTACCAAT 4858
Db 863 -----LeuLysSerProThrMetHisGlnValGlnSerPrometLeuGlySerProSer 880
QY 4859 ATACAGATGAGGTAATATATTAAATTTATTATTAAAGTTTGTGTTAATTATCTTC 4918
Db 881 GlyAsnLeu----- 883
QY 4919 TTTTCAGCGTCAAGCATCAGCGCAAGGTGATCCGTACAAATTAGTCGC----- 4969
Db 884 -----LysSerProGlnThrProSerGlnLeuAlaGlyMetLeuAla 897
QY 4970 -----GCTCCGATAATATCCGCTAAATCCCAATAGTGGCAATCGCGCC----- 5015
Db 898 GlyProAlaAlaAlaAlaSerIleLysSerProProValLeuGly-SerAlaAlaAlaAla 917
QY 5016 -----ACCAACAAGA 5026
Db 917 rProValHisLeuLysSerProSerLeuProAlaProSerProGlyTrpThrSerSerPr 937
QY 5027 TGACCCAAACTTCGATCCAACTCTTCTTTGGCACAATGTCCCAACAACATAACAGTT 5086
Db 937 OlyProProLeuGlnSer-----ProGlyIleProProAsnHisIlySAl 952
QY 5087 GCGTGTCCAGCATGGGTAGTCCAGCCGGAAGTGTGTATGACGATGAGGAGTCCG- 5145
Db 952 aProLeu-ThrMetAlaSerProAlaMetLeuGlyAsnValGlnSerGlyGlyProProP 972
QY 5146 -----GGACCGTCGACATCAATATTGAG----- 5169
Db 972 roProThrAlaSerGlnProAlaSerValAsnIleProGlySerLeuProSerSerThrp 992
QY 5169 ----- 5169
Db 992 roTyThrMetProProGluProThrLeuSerGlnAsnProLeuSerIleMetMetSera 1012
QY 5170 -----CATGAATAATTCCGGAC 5188
Db 1012 rGmetSerLysPheAlaMetProSerSerThrProLeuTyHisAspAlaIleLysThrV 1032
QY 5189 TAGATGATCAGGAATAGATACC-----ATAATCAAAATAACTGTCAATCAATGAATG 5242
Db 1032 aAlaSerSerAspAspAspSerProProAlaArgSerProAsnLeuProSerMetAsn- 1051
QY 5243 TCGTAATGAACCTCAATGGGTCCCGGAATGCTGAATCCTTAAATGTGCGTAGACGGGTC 5302
Db 1052 --AsnMetProGlyMetGlyIleAsnThrGlnAsnProArgIle-----SerGlyP 1068
QY 5303 CAAAT----- 5307
Db 1068 roAsnProValValPrometProThrLeuSerPrometGlyMetThrGlnProLeuSerH 1088
QY 5308 -----GACCCGCTGGCTTTAATCCTAATTTCCCAATGTTGATTA 5350
Db 1088 iAsSerAsnGlnMetProSerProAsnAlaValGlyProAsnIleProProHisIlyVal- 1107
QY 5351 GAGAGAATTCATAGGCTGTGGCTGTGGCTCAGCAAACTCTTCAAACCTTCAAGGGGTTG 5410
Db 1108 -----PrometGlyProGlyLeuMetSerHisAsnProIleMetGlyHisIlySerG 1125
QY 5411 TTCACCTGTGTCAGAAATGATGGTCCGAATGCCAGTCAATTTGGTTCGAATTTCAATC 5470
Db 1125 lngluProPrometValProGlnGlyArgMet-----GlyPheProGlnGlyPheProP 1143
QY 5471 CGAATATTCAAGTAAGCGAGTACCCCAACACATACAATACATGCGAGTAAGGCGAC 5530
Db 1143 ro-----ValGlnSerProProGlnGlnValP 1152
QY 5531 AGAACGCCAACACAATAACAACATGAGAGCTAATAATGTGCGAATGCCACCTAGTCTGG 5590

Db 1152 roPheProHISaNGlyProSerGlyGlyGlySer-----PheProGlyGlyMetG 1170
OY 5591 AATTTTGcAGAGGTACGCTAACCCCTCAATGGGTGCTAGGCAATGGTCCCAATAT 5650
Db 1170 lYPhe-----ProGlyGlyGlyProLeuGlyAArgProSerAsnLeu- 1183
OY 5651 GCCCACCATCAGCCAGCGAGCTACTCCTGGAATGCCAGATTGATGGCGGACGAGAG 5710
Db 1184 --ProGlnSerSerAlaSpAlaAlaLeuCysLysProGlyGlyProGlyGlyProAspS 1203
OY 5711 CCGGAGGTATGCTAATGAATTCT----- 5733
Db 1203 erPheThrValLeuGlyAsnSerMetProSerValPheThrAspProAspLeuGlnGluV 1223
OY 5733 ----- 5733
Db 1223 allleArgProGlyAlaThrGlyIleProGluPheAspLeuSerArgIleIleProSerg 1243
OY 5734 -----TCCGAGAGCAACACACAGACAGATCA 5761
Db 1243 lulysProSerGlnThrLeuGlnTyrPheProArgGlyGluValProGlyArgLysGlnP 1263
OY 5762 CAACAATCCTGGGCGAAGCAATGTATTACTTTTCAGAAATTGCAATCAATGTCTA 5821
Db 1263 roGlnGlyProGlyPro-----GlyPheSerHisMetGln-----GlyM 1276
OY 5822 TTGTTGACGAGAGGT-----GGATTACCCGCGCATGACGATCAATGA 5866
Db 1276 etMetGlyGluGlnAlaProAArgMetGlyLeuAlaLeuProGlyMetGlyGlyProGlyP 1296
OY 5867 ATATTGTCACCATCTATGATAAGGGCATGCGTCCA-----CATGCCA 5911
Db 1296 roValGlyThrProAspIleProLeuGlyThrAlaProSerMetProGlyHisAsnProm 1316
OY 5912 TGCGGCGCAAT-----GTAATGGGTGCGCGGATGCCACCCGTTACAGGCAATTCACT 5965
Db 1316 etArgProProAlaPheLeuGlnGlnGlyMetMetGlyProHisHisArgMetMetSerp 1336
OY 5966 TTGCACAGTCATCGGATGTAATTGACTGTGTCGGGATCCGTCATCATTTTTCACTAACG 6025
Db 1336 roAlaGlnSerThr-----MetProGlyGlnProThrLeuMetSerAsnProA 1352
OY 6026 CTTCCTGCAACAGCGCTGACCAACATGTTTGATCAGACACAAGGCCAATCAGCCTA 6085
Db 1352 laAlaAla----- 1354
OY 6086 AGACACAACACATAAAGACATACCTAGTGAATGTGTCAAAACCAATCGGACTTGCAg 6145
Db 1355 -----ValGlyMetIlePro---GlyLysAspAArgGlyProAlaGlyLeuTyrT 1370
OY 6146 TGGCACAAGGCGAGATCCCACTGCATGGGCAAGACATGCCGACAGGTCAGTCTTAATG 6205
Db 1370 hrHisProGlyProValGlySerProGlyMetMetSerMet---GlnGlyMetMetG 1389
OY 6206 GACCTACTAATAATAATT-----ATGTCAACTGCCGGAAGTGTCACTAGTCTA 6253
Db 1389 lYPProGlnGlnAsnIleMetIleProProGlnMetAArgProArgGly--MetAlaAla 1408
OY 6254 CTAACGGTGTCTTGCGATCAATTTCGTAGTCCCTCTTCT 6294
Db 1408 spValGlyMetGlyGlyPheSer---GlnGlyProGlyAsn 1420

RESULT 12

ADJ71903
ID ADJ71903 standard; protein; 1426 AA.
XX
AC ADJ71903;
XX
DT 20-MAY-2004 (first entry)
XX
DE Human Lgs/Bcl9 polypeptide.
XX

KW Human; legless; lgs; cell differentiation disorder;
KW cell proliferation disorder; cancer; Wnt pathway; medulloblastoma; colon;
KW breast; head; neck; brain; thyroid; skin; blood disease;
KW tissue regeneration; tissue repair; cytostatic; lgs/Bcl9.
OS Homo sapiens.
XX
PN US2004038901-A1.
PD 26-FEB-2004.
XX
PF 22-SEP-2003; 2003US-00664859.
XX
PR 28-JUL-2000; 2000US-0221502P.
PR 27-JUL-2001; 2001US-00915543.
XX
PA (UYZU-) UNIV ZURICH.
XX
PI Basler K, Brunner E, Froesch B, Kramps T, Peter O;
XX
DR WPI; 2004-203288/19.
DR N-PSDB; ADJ71902.
XX
PT Novel polypeptide sharing one or more homologue amino acid domains with
PT Legless protein being functional homologue of Legless, useful for
PT diagnosing disorders of cell fate.
XX
PS Example 2; SEQ ID NO 15; 62pp; English.
XX
CC The invention relates to a polypeptide sharing one or more homologous
CC amino acid domains with a legless (lgs) protein and is therefore a
CC functional homologue of lgs. The invention also relates to a nucleotide
CC sequence encoding a protein present in invertebrate and/or vertebrate
CC organisms, the nucleotide sequence encoding a protein comprising a
CC positive function in a regulatory pathway and the use of the polypeptide
CC for the isolation of lgs-binding proteins by carrying out an assay chosen
CC from an in vitro binding assay with such a peptide or a co-
CC immunoprecipitation from vertebrate or invertebrate cell lysates or a
CC mammalian or yeast two hybrid assay. The polypeptide and polynucleotide
CC are useful for treating disorders of cell fate, which involves
CC administering therapeutic compounds chosen from invertebrate and
CC vertebrate lgs protein homologues or fragments, antibodies, antibody
CC fragments, lgs antisense DNA, lgs antisense RNA, lgs double-stranded RNA,
CC small peptides or chemical and natural compounds being capable of
CC interfering with lgs function, synthesis and degradation. The disorders
CC are related to cell differentiation or cell proliferation. The compound
CC is administered to treat a cancerous condition by preventing progression
CC from a pre-neoplastic or non-malignant condition to a neoplastic or
CC malignant state. The cancerous condition is characterised by over-
CC stimulation of the Wnt pathway and is medulloblastoma or cancer of the
CC colon, breast, head and neck, brain, thyroid or skin. The therapeutic
CC compound may also be administered to a blood disease to promote tissue
CC regeneration and repair. This sequence represents the human lgs/Bcl9
XX polypeptide of the invention.
SQ Sequence 1426 AA;
XX
Alignment Scores:
Pred. No.: 4.69e-17 Length: 1426
Score: 312.00 Matches: 313
Percent Similarity: 30.34% Conservative: 177
Best Local Similarity: 19.38% Mismatches: 519
Query Match: 2.59% Indels: 607
DB: 8 Gaps: 73
US-10-664-859-1 (1-6909) x ADJ71903 (1-1426)
OY 2397 AAAATTGAAGA--ATTCAAGACAGTACCGAAGGATCGTCTTGACA 2453
Db 98 LysArgGluArgSerIleSerAlaAspSerPheAspGlnArgAspProGlyThr----- 115
OY 2454 ATGATATATGACGAATGAGCATGGAAGGCTGCAATCACTGATCCGATTTATCAAT 2513

Db 116 ---ProAsnAspAspSerAspIlelyeGluCysAsnSerAlaAspHisIlelySerGln 134
QY 2514 GAATCTTTAAATAATCCTCGCAATTTGCGCATATTAGTAAGCGGAGTAGGACCAATACC 2573
Db 135 AspSerGlnHisThrProHisSerMetThr----- 144
QY 2574 GGAATCGGAGTTGGAGCGGGGAGCGGAAATTATTAGTCTGCCACGCCAATGGAATCTCC 2633
Db 145 -----ProSerAsnAlaThrAlaProArgSerSerThrProSer 157
QY 2634 TCGGTAGCAGTAATTGTTGGATTACATGCAACAGCAAAATCAC-----ATA 2681
Db 158 HisGlyGlnThrThrAlaThrGluProThrProAlaGlnLysThrProAlaLysValVal 177
QY 2682 TTGCTGTTTTCACTCAGCTGGCCAAAGGGCCGGAATCAGTTTAAAGCGTCAATTT 2741
Db 178 TyrValPheSerThrGluMetAlaAsnLysAlaAlaGluAlaValLeuLysGlyGlnVal 197
QY 2742 CAACTATATTGCGTATCACTGCACCTGCCTCTACAAAAGCTTCTGGAAGACTTT 2801
Db 198 GluThrIleValSerPheHis----- 204
QY 2802 TTTATGAAAAACCCTTTAAAGATTACAGTTACAGCGGACAAATTCGTCGTATGCCA 2861
Db 205 ---IleGlnAsn---IleSerAsnAsnLysThrGluArgSerThrAla-----Pro 219
QY 2862 TGGATAGGATGGGGCAGGTGGACTTAATCTCTTAATCCTGTAGCCAAATAACACAA 2921
Db 220 LeuAsnThrGlnIleSerAlaLeuArgAsnAspProLysProLeu-----Pro 235
QY 2922 CAGACGCCACATACAAAGACCGTAGCGCTATTGAAACCCCAATTCATCAACATGAAAC 2981
Db 236 GlnGlnProProAlaProAla-----AsnGlnAspGlnAsn 247
QY 2982 AGCAACGTAAGTACTGTAAGCGCGCCTAGCACTCTTT----- 3020
Db 248 SerSerGlnAsnThrArgLeuGlnProThrProProIleProAlaProAlaProLysPro 267
QY 3021 -----GTCGACCAGTCTGATCCTATGGGCAACGAACTGAATTGATG 3062
Db 268 AlaAlaProProArgProLeuAspArgLysSerPro--GlyValGlnAsnLysLeuIle 286
QY 3063 TGCTGGGAAGCGGATCCTCAAAACACCAGT-----AGTCTGACAAAAAC 3107
Db 287 ProSerValGlySerProAlaSerSerThrProLeuProProAspGlyThrGlyProAsn 306
QY 3108 TCACGAATCATGTAGACAGTATCAGTATCCACGAGTCACAGCAATAAAGATACTG 3167
Db 307 SerThrProAsnAsnArgAlaValThrProValSerGlnGlySerAsnSerSerSerAla 326
QY 3168 GAAGCAGCTGGCGTTGATTGGGACAGGTGCACAAAAGAGCGATCCTGGCTGACAACT 3227
Db 327 AspProLysAlaProProProProProValSerSerGlyLysProProThrLeu---Gly 345
QY 3228 GAAACACACATTGTATCACTGCAAGAGATTAAAGTTCCAGACGAAACCTTACACCACAA 3287
Db 346 GluAsn-----ProAsp---GlyLeuSerGlnGlu 354
QY 3288 CAGCGGCAACATCGGGAAGAACAGTGGCAAAATAAAAAATGAATCAATTTCTTTT 3347
Db 355 GlnLeuGlnHisArgGluArgSerLeuGlnThrLeuArgAspIleGlnArgMetLeuPhe 374
QY 3348 CCTGAAATGAGAAATTCAGTAGAGCTAATGTAGCTCACAGATACAAAAATTCCAGGA 3407
Db 375 ProAspGluLysGluPheThrGlyAlaGln----- 384
QY 3408 GATTTAATGATGGGATGTCGGGTGCGGAGCGGATCTATTATAAATCCGACGATGCGA 3467
Db 385 -----SerGlyGlyProGlnGlnAsnPro----- 392
QY 3468 CAACTGCATATGCCAGGTAACGCCAAATCGAGCTTATCGCGACAACTTCAGACTT 3527
Db 392 ----- 392

QY 3528 TCGGAAGATGTAATGCATCCAGGGGATGTTATATCAGATATGGTGCCGTAATAGATGT 3587
Db 393 -----GlyValLeuAspGlyPro 398
QY 3588 AATAATAATCAAAAAACAGTGTGCAATGTGATCTGGAGTAGGTGTGTCACTGGAACA 3647
Db 399 GlnLysLysProGluGlyProIleGlnAla----- 408
QY 3648 ACTGCAGCTGGAGTAATGTCAATATGCATTGCTCAAGCTCCGGCCCCGAATGGCAAT 3707
Db 409 -----MetMetAlaGlnSerGlnSerLeuGlyLysGly--- 419
QY 3708 ATGATGGGAAGCTCTACGGATATGCTAGCCCTGTTGGCAACACAAAGCTGCAACGTCAATC 3767
Db 420 --ProGlyProArgThrAspValGlyAlaProPheGlyProGlnGlyHisArgAspVal 438
QY 3768 GGAACGCCCCAGAT--ATGTCCTAAGGAAGTTTAAATCAAGATAGCCGAAC----- 3818
Db 439 ProPheSerProAspGluMetValProProSerMetAsnSerGlnSerGlyThrIleGly 458
QY 3819 -----CATTCACATCAAGGGGAGTTGCTCAAAATGAGTGTGGAAGATTCAAACATCAA 3872
Db 459 ProAspHisLeuAspHisMetThrProGluGlnIleAlaTrpLeuLysLeuGlnGlnGlu 478
QY 3873 TTTTTCGAAGAAGCGCCTCAAGGGGGCAAGCCACAGACAAAGTCACTGGAACCTGTAGTACCA 3932
Db 479 PheTyrgluGluLysArgArg-----LysGlnGlnGlnVal-----ValValGln 493
QY 3933 CAA-----CAGCAAAACCCCTTCTGATCTGTGGAAC 3965
Db 494 GlnCysSerLeuGlnAspMetMetValHisGlnHisGlyProAlaArgGly----- 509
QY 3966 TCGTTAAACAACAGGTGCGACCCCTGCAAGGTCCACCTCTCTTACCAC----- 4016
Db 510 -----ValValArgGlyProProProProTyrglnMetThrPro 522
QY 4017 -----TCCATCCAGAGATCTGCG 4034
Db 523 SerGluGlyTrpAlaProGlyGlyThrGluProPheSerAspGlyIleAsnMetProHis 542
QY 4035 TCAGTACCAATAGCCACTCAATCGCCCAATCCCTCGAGTCCAAACAATCTA----- 4085
Db 543 SerLeuProProArgGlyMetAlaProHisProAsnMetProGlySerGlnMetArgLeu 562
QY 4086 -----TCTCTCCCGTCACCGCGG 4103
Db 563 ProGlyPheAlaGlyMetIleAsnSerGluMetGluGlyProAsnValProAsnProAla 582
QY 4104 ACAACCGCAGCAGTCATGGGATG-----CCGACCAACTCTCTAGCATG----- 4148
Db 583 SerArgProGlyLeuSerGlyValSerTrpProAspAspValProLysIleProAspGly 602
QY 4149 -----GATGGAACAGGATCATTTACTGGA----- 4172
Db 603 ArgAsnPheProProGlyGlnGlyIlePheSerGlyProGlyArgGlyLysArgPhePro 622
QY 4173 -----TCTGTTCCGCAAGCTAATACTTCAGACGGTTCAGGACAGC 4211
Db 623 AsnProGlnGlyLeuSerGluGluMetPheGlnGlnGlnLeuAlaGluLysGlnLeuGly 642
QY 4212 ACAACA-----ACAGTGTCTCGCAACAAGAAGACTGTTTCAGGACAGACCCCATCG 4265
Db 643 LeuProProGlyMetAlaMetGluGlyIleArgProSerMetGluMetAsnArgMetIle 662
QY 4266 CCGTCAAAATCAAAATCGTAGTAGAAATAACCGGATCGTCAAGCCTTCTTACGCATACTTA 4325
Db 663 ProGlySerGln--ArgHisMetGluProGlyAsnAsnProIlePheProArgIlePro 681
QY 4326 AGCAGCAACCAAGTACCCCTTATCTCATCTATCCCAAG----- 4367
Db 682 ValGluGlyProLeuSerProSerArgGlyAspPheProLysGlyIleProProGlnMet 701

| | | | |
|----|------|--|------|
| QY | 4368 | -----GAATTGAGTCTTTCGGTCAGTCCCTGCTGCTATGTATATTGTTT | 4415 |
| Db | 702 | GlyProGlyArgGluLeuGlu--PheGlyMetValProSerGlyMet----- | 716 |
| QY | 4416 | AATTTTTAAAGACAATAATCAATATGCAATTGCGTTAAATTAAGTTATATATTACATAA | 4475 |
| Db | 717 | -----LysGlyAspValAsnLeuAsn--ValAsn----- | 725 |
| QY | 4476 | CTCCGAATTTGATAGAAAAATCAGGAATAGAAAAATAAATTATTTTCCGACCG-- | 4532 |
| Db | 726 | MetGlySerAsnSerGlnMetIleProGlnLysMetArgGluAlaGlyProGlu | 745 |
| QY | 4532 | ----- | 4532 |
| Db | 746 | GluMetLeuLysLeuArgProGlyGlySerAspMetLeuProAlaGlnGlnLysMetVal | 765 |
| QY | 4533 | ----- | 4547 |
| Db | 766 | ProLeuProPheGlyGluHisProGlnGlnGlnLutyrGlyMetGlyProArgProPheLeu | 785 |
| QY | 4548 | -----AATCCAATTCTGAGTGATGTTAGAGATAA----- | 4579 |
| Db | 786 | PrometSerGlnGlyProGlySerAsnSerGlyLeuArgAsnLeuArgGluProIleGly | 805 |
| QY | 4580 | -----TCTACTATTAAATTAACAACAGAAAAATTCATATCCGTTAATTGAAT | 4627 |
| Db | 806 | ProAspGlnArgThrAsnSerArgLeuSerHisMetProProLeuProLeu----- | 822 |
| QY | 4628 | CACTATTGTTTAAATAAGAAATTAATAATATGTTTATTATTAATTCTACAGGTGATAC | 4687 |
| Db | 823 | -----AsnProSerSerAsn | 827 |
| QY | 4688 | ATGAAAAGTAGGCCGACCAAGCCACAGGGTCAGCGGTCAACGATAATAGTCTAATAGAG | 4747 |
| Db | 828 | ProThrSerLeuAsnThrAlaProProValGlnArg-----GlyLeuGlyArg | 843 |
| QY | 4748 | GCAATATAAGATGTACGATTGCTGCATCC-----AGTCTGTTTAAACCGCAT | 4798 |
| Db | 844 | LysProLeuAspIleSerValAlaGlySerGlnValHisSerProGlyIleAsnPro--- | 862 |
| QY | 4799 | CCACATATGCAAGCAATTCAAAATTCAGCATTAACGCGCTATAAATGGCTTACCAAT | 4858 |
| Db | 863 | -----LeuLysSerProThrMetHisGlnValGlnSerPrometLeuGlySerProSer | 880 |
| QY | 4859 | ATACAGATGAGGTAATATTTAATATTATTATTAAAGTTTGTGTTAATTATCTTC | 4918 |
| Db | 881 | GlyAsnLeu----- | 883 |
| QY | 4919 | TTTTTCAGCGTCAAGCATCAGCCGAAGGTGATCCGTACAATTAGTCGCG-- | 4969 |
| Db | 884 | -----LysSerProGlnThrProSerGlnLeuAlaGlyMetLeuAla | 897 |
| QY | 4970 | -----GCTCCGATTAATATTCGCTAAATCCCAATAGTGGCAATCGCGCGC---- | 5015 |
| Db | 898 | GlyProAlaAlaAlaAlaSerIleLysSerProProValIleuGly-SerAlaAlaAla | 917 |
| QY | 5016 | -----ACCAAAACAGA | 5026 |
| Db | 917 | rProValHisLeuLysSerProSerLeuProAlaProSerProGlyTrpThrSerSerPr | 937 |
| QY | 5027 | TGACCCAAACTTCGATCCAATCTCTTTTGGACAATAATGCCCAACAATAACAAGTT | 5086 |
| Db | 937 | OlfsProProLeuGlnSer-----ProGlyIleProProAsnHisIlyVal | 952 |
| QY | 5087 | GCGTGTCCAGCATGGGTAGTCCAGCCGGAAGTGGTGGTATGACGATGATGGGGGTCCG- | 5145 |
| Db | 952 | aProLeu-ThrMetAlaSerProAlaMetLeuGlyAsnValGluSerGlyGlyProProP | 972 |
| QY | 5146 | -----GGACCGTCCGACATCAATATTGAG----- | 5169 |
| Db | 972 | roProThrAlaSerGlnProAlaSerValAsnIleProGlySerLeuProSerSerThrP | 992 |
| QY | 5169 | ----- | 5169 |

| | | | |
|----|------|---|------|
| Db | 992 | roTyrThrMetProProGluProThrLeuSerGlnAsnProLeuSerIleMetMetSera | 1012 |
| QY | 5170 | -----CATGAAATAATTCCGGAC | 5188 |
| Db | 1012 | rgMetSerLysPheAlaMetProSerSerThrProLeuTyRHisAspAlaIleLysThrV | 1032 |
| QY | 5189 | TAGATGATCAGGAATAGATACC-----ATAAATCAAAATAACTGTCAATGCAATGAATG | 5242 |
| Db | 1032 | alAlaSerSerAspAspAspSerProProAlaArgSerProAsnLeuProSerMetAsn- | 1051 |
| QY | 5243 | TCGTAATGAACCTCAATGGCTCCCGAATGCTGAATCTAAATATGTCGTAGCAGCGCGTC | 5302 |
| Db | 1052 | --AsnMetProGlyMetGlyIleAsnThrGlnAsnProArgIle-----SerGlyP | 1068 |
| QY | 5303 | CAAAAT----- | 5307 |
| Db | 1068 | roAsnProValProMetProThrLeuSerProMetGlyMetThrGlnProLeuSerH | 1088 |
| QY | 5308 | -----GGACCGCCTGCTTTAAATCTTAATCCCGCAATGGTGGATTA | 5350 |
| Db | 1088 | isSerAsnGlnMetProSerProAsnAlaValGlyProAsnIleProProHisGlyVal- | 1107 |
| QY | 5351 | GAGAGAATTCATAGGGTCTGGCTGTGGCTCAGCAACTCTTCAACTTTCAAGGGGTG | 5410 |
| Db | 1108 | -----PrometGlyProGlyLeuMetSerHisAsnProIleMetGlyHisGlySerg | 1125 |
| QY | 5411 | TTCCACCTGTGCGCAGATGATGGTGGAAATGCCAGTCAATTTTGTTCGAATTTCAATC | 5470 |
| Db | 1125 | IngIuProPrometValProGlnGlyArgMet-----GlyPheProGlnGlyPheProP | 1143 |
| QY | 5471 | CGAATATTTCAGTAAAGCGAGTACCCCAACACATACATACATGCCAGTAAGGCGAC | 5530 |
| Db | 1143 | ro-----ValGlnSerProProGlnGlnValP | 1152 |
| QY | 5531 | AGAACGCCAACACAATAACAACATGAGCTAATAATGTGCGAATGCCACCTAGTCTGG | 5590 |
| Db | 1152 | ropheProHisAsnGlyProSergIlyGlnGlySer-----PheProGlyGlyMetG | 1170 |
| QY | 5591 | AATTTTTCAGAGGTAACGCTAACCTCAAAATGGTGCTGTAGCAATGGTTCGCCAATAT | 5650 |
| Db | 1170 | lyPhe-----ProGlyGlnGlyProLeuGlyArgProSerAsnLeu- | 1183 |
| QY | 5651 | GCCCACCATCAGCCAGCGAGCTACTCCTGGAATGCCAGATTGATGCGCGACAGAG | 5710 |
| Db | 1184 | --ProGlnSerSerAlaAspAlaAlaLeuCysLysProGlyGlyProGlyProAsps | 1203 |
| QY | 5711 | CCGAGGTATGCTAATGAATTCT----- | 5733 |
| Db | 1203 | erPheThrValLeuGlyAsnSerMetProSerValPheThrAspProAspLeuGlnIuv | 1223 |
| QY | 5733 | ----- | 5733 |
| Db | 1223 | alIleArgProGlyAlaThrGlyIleProGluPheAspLeuSerArgIleIleProSerg | 1243 |
| QY | 5734 | -----TCCGAGAGCAACACAGAGATCA | 5761 |
| Db | 1243 | IuLysProSergIlnThrLeuGlnTyRpheProArgIlyGlyValProGlyArgLysGlnP | 1263 |
| QY | 5762 | CAAAACATCCTGGGGCAGCAATGTATTACTTCTTTCAGAAATTCGAATCAATGTCTA | 5821 |
| Db | 1263 | roGlnGlyProGlyPro-----GlyPheSerHisMetGln-----GlyM | 1276 |
| QY | 5822 | TTGTGTGACGAAGAGGT-----GGAATTAACCGGCCATGACGATCAATGA | 5866 |
| Db | 1276 | etMetGlyGlnGlnAlaProArgMetGlyLeuAlaLeuProGlyMetGlyGlyProGlyP | 1296 |
| QY | 5867 | ATATTGCTCAACCATCTATGATATAAGGGCGATCGCTCA-----CATGCCA | 5911 |
| Db | 1296 | roValGlyThrProAspIleProLeuGlyThrAlaProSerMetProGlyHisAsnProm | 1316 |
| QY | 5912 | TGCGGCCAAAT-----GTAATGGGTGCGCGGATGCCACCCGTTAACAGGCAATTCAAGT | 5965 |

Db 1316 etArgProProAlaPheLeuGlnGlyMetMetGlyProHisHisArgMetMetSerp 1336
QY 5966 TTGCACAGTCATCGGATGGTATTGACTGTGCGGGATCCGTCATCATTTTCACTAACG 6025
Db 1336 roAlaGlnSerThr-----MetProGlnGlnProThrLeuMetSerAsnProA 1352
QY 6026 CTTCCTGCAACAGCGCTGGACCACACATGTTGGATCAGACAAAGGCCAATCAGCCTA 6085
Db 1352 laAlaAla----- 1354
QY 6086 AGACACAACACATAAAGACATACCTAGTGAATGTGCAAAACCAATCGGACTTGCAg 6145
Db 1355 -----ValGlyMetIlePro--GlyLysAspArgGlyProAlaGlyLeuTyrT 1370
QY 6146 TGGCACAAGGCGAGATCCCACTGCATGGGCAAGACATGCCGAGGTCAGTCTTAATTG 6205
Db 1370 hrHisProGlyProValGlySerProGlyMetMetMetSerMet---GlnGlyMetMetG 1389
QY 6206 GACCTACTATAATAATTA-----ATGCAACTGCCGGAAGTGCAGTGCCTA 6253
Db 1389 lYProGlnGlnAsnIleMetIleProProGlnMetArgProArgGly--MetAlaAlaA 1408
QY 6254 CTAAACGGTGTCTCTGCGCATCAATTGCGTAGTCCCTCTTCT 6294
Db 1408 spValGlyMetGlyGlyPheSer---GlnGlyProGlyAsn 1420

RESULT 13

AAU78460
ID AAU78460 standard; protein; 1494 AA.

AC AAU78460;

DT 02-JUL-2002 (first entry)

DE Mouse beta-catenin nuclear localised protein.

KM Mouse; beta-catenin nuclear localised protein; cancer; gene therapy; EST;

KW expressed sequence tag.

OS Mus musculus.

PN W0200224738-A1.

PD 28-MAR-2002.

PF 19-SEP-2001; 2001WO-JP008140.

PR 22-SEP-2000; 2000JP-00287876.

PA (KYOW) KYOWA HAKKO KOGYO KK.

PI Akiyama T, Adachi S;

DR WPI; 2002-330014/36.

DR N-PSDB; ABK47631.

PT New beta-catenin nuclear localized protein for diagnosis and treatment of

PT diseases associated with nuclear localization of beta-catenin e.g.

PT cancer.

PS Claim 1; Page 81-88; 113pp; Japanese.

CC The invention relates to a beta-catenin nuclear localised protein and DNA

CC encoding the protein. The protein and encoding DNA are applicable in

CC diagnosis and treatment of diseases associated with nuclear localisation

CC of beta-catenin e.g. cancer, including gene therapy. The present sequence

CC represents the amino acid sequence of mouse beta-catenin nuclear

CC localised protein

SQ Sequence 1494 AA;

Alignment Scores:

Pred. No.:

1.03e-15

Length:

1494

Score: 296.50 Matches: 308
Percent Similarity: 30.65% Conservative: 164
Best Local Similarity: 20.00% Mismatches: 534
Query Match: 2.46% Indels: 535
DB: 5 Gaps: 79

US-10-664-859-1 (1-6909) x AAU78460 (1-1494)

QY 2484 TGCAATCAGTTGAATCCGATTTTATCAATGAATCTTTAAATAATCCTGCAATTCGAGC 2543
Db 179 CysAsnValAlaAspProAlaMetValThrProGlnLeuGlyProGlyGlnThrAlaGln 198
QY 2544 ATATTAGTAAGCGGAGTA-----GGACCAATACCCGGAATCGGA 2582
Db 199 LeuProLeuSerGluSerSerAlaProGlyProGlnHisGlyProGlnProGlyLeu-- 217
QY 2583 GTTGAGCGGGGACGGGAATTATTTGACTGCCAACGCCAATGGAATCTCTCGGTAAGC 2642
Db 218 -----ArgProAspValProGlyGly 225
QY 2643 AGTAATTGTTGATTACATGCAACAGCAAAATCACATATTCGTTTCAACTCAGCTG 2702
Db 226 GlyGlyGlyValProGlyLysProProSerGlnPheValTyrValPheThrThrHisLeu 245
QY 2703 GCCAACAAGGGCGGCAATCAGTTTAAAGCGGTCAATTTCAACTATTATTCGCTATCAC 2762
Db 246 AlaAsnThrAlaAlaGluValLeuGlnGlyArgAlaGluSerIleLeuAlaTyrHis 265
QY 2763 TGCACTCAGCCTGCTACAAAAGCTTCCTGGAAGACTTTTATGAAAAAACCTTTAAG 2822
Db 265 ----- 265
QY 2823 ATTACAAGTTACAGCGCACAAATCCGTCGTATGCCATGATAGCATGGGAGGTT 2882
Db 266 -----GlnGlnAsnValProArgAlaLysLeuAspGlnAla 277
QY 2883 GGACTAACTCCTCTATCTGTAACCCAAATAACACACAGCACCA-----CAT 2933
Db 278 ProLysValProProThrPro-----GluProLeuProLeuAsn 290
QY 2934 ACAAGACCGTAGCGCTATTGAAACCCCAATTCATCAACATGAAACAGCAACGTAAGT 2993
Db 291 ThrProSerAlaGlyThrProGlnSerGlnProProProLeuProProProAla 310
QY 2994 ACTGTAAGCGCGCTAGCAACTCTTTTGTGCAGCAGTGTGATCTATGGGCAACGAACT 3053
Db 311 ProGlySerAlaProPro--AlaLeuProProGluGlyProProGluAspThrSerGln 329
QY 3054 GAATTGATGTGCTGGGAAGCGGATCCTCAACACACAGTAGGTGACAAAACTCACGA 3113
Db 330 AspLeuAlaProAsnSerValGlyAlaAlaSerThrGlyGlyThrGlyGlyThrHis 349
QY 3114 AATCATGTAGACAGTATCAGTACATCCAGCAGTCAAGGCAATAAAGATACTGAAGCA 3173
Db 350 ProAsnThrProThrAlaAlaThrAlaAsn----- 359
QY 3174 GCTGGCGTTGATTGGGACAGGTCAAAAGGAAGCGATCCTGCGCTGACAACCTGAAGAAC 3233
Db 360 -----AsnProLeuProProGlyGlyAspProGlySerAlaProGlySer 374
QY 3234 AACATTGTATCCTGCAAGGAGTTAAGTTCCA----- 3266
Db 375 -----AlaLeuLeuGlyGluAlaThrProThrGlyAsnGlyGlnArgAsnLeuVal 391
QY 3267 ---GACGAAAACCTTACACCACCAAGCGGCAACATCGGGAAGAACAGTTGGCAAAATA 3323
Db 392 GlySerGluGlyLeuSerLysGluGlnLeuGluHisArgGluArgSerLeuGlnThrLeu 411
QY 3324 AAAAAAATGAATCAATTTCTTTCTGAAAATGAGAATTGAGATTGAGAGCTAATGTAAAGC 3383
Db 412 ArgAspIleGluArgLeuLeuLeuArgSerGlyGluThrGlu----- 425
QY 3384 TCACAGATTAACAAAAATTCCAGAGAGATTTAATGATGGGATGTGGGTGGCGAGCGGA 3443

| | | | | | | |
|----|------|---|--|---|-------|------|
| Db | 426 | ----- | Proheuleuysg | ProProglglyalagly | --- | 436 |
| QY | 3444 | TCTATTATAATCCGACGATCGACAACTGCATATGCCAGGTAAACGCCAATCCGAGCTC | | | | 3503 |
| Db | 437 | ----- | Glulglyl | ProProAlaGlnAla | | 444 |
| QY | 3504 | TTATCGGCGCAAGTTCAGGACTTTCGGAAGATGTATGCATCCAGGGATGTTATATCA | | | | 3563 |
| Db | 445 | ProSerAlaAlaGlnProProProSerAlaPro | ----- | ProGlyGlyLeu | ----- | 459 |
| QY | 3564 | GATATGGTGCCGTAATAGATGTAATAATCAAAAAACAGTGTCAATGTGATCT | | | | 3623 |
| Db | 460 | ----- | LysLsTyrglulProleuGln | ----- | | 467 |
| QY | 3624 | GGAGTAGTGTGTCTACTGGAACAACACTGCAGCTGGAGTAATGTCAATATGCAATTGCTCA | | | | 3683 |
| Db | 468 | ----- | SerMetIleSerGlnThrGlnSerleuGlyProProleuGlnHis | --- | Glu | 484 |
| QY | 3684 | AGCTCCGGCGCCCGAATGCGCAATATGATGGGAAGCTTACGGATATGCTAGCCTCGTTT | | | | 3743 |
| Db | 485 | ValProGlyHisProGlnGlyGlyAspMetGlyGlnMetAsnMetMet | ----- | | | 501 |
| QY | 3744 | GGCAACACAAGCTGCAACGTCATCGGAACGGCCCCAGATATGCTTAAGGAAGTTTAAAT | | | | 3803 |
| Db | 502 | ----- | MetGlnArgleuGly | | | 506 |
| QY | 3804 | CAAGATAGCCGAACCCATTCATCATCAAGGGGAGTTGCTCAATGAGTGTGCAAGATT | | | | 3863 |
| Db | 507 | GlnAspSerLeuThrProGlu | ----- | GlnValAlaTrpArgLysLeu | | 520 |
| QY | 3864 | CAACATCAATTTTTCGAAGAAGC | ----- | CTCAAGGGGAGC | | 3899 |
| Db | 521 | GlnGluGluTyrrTyrgluGluLysArgArgLysgluGlnIleGlyLeuHisglYgLy | | | | 540 |
| QY | 3900 | AAGCCAGACAAGTCACTGGAACCTAGTACCACAACAACAACCCCTTCTGATCTGCT | | | | 3959 |
| Db | 541 | ArgPro | ----- | LeuGlnAspMetValGlyMetGly | | 550 |
| QY | 3960 | GGAACCTCGTTAAACAACAGGTGCGACCCCTGCAAGSTCCACCTCTCTTACCACCTCC | | | | 4019 |
| Db | 551 | Gly | ----- | MetMetGlyArgGlyProProProProIleHisSer | | 563 |
| QY | 4020 | ----- | ATCCAGAGATCTGCGTCACTACCAATAGCCACT | | | 4052 |
| Db | 564 | LysProGlyAspGlnCysAlaProGlyMetGlyAlaGlnLeuArgGlyProMetAspVal | | | | 583 |
| QY | 4053 | CAATCGCCC | ----- | AATCCCTCGAGTCCAAACAATCTATCTCTCCCGTACCCGCG | | 4103 |
| Db | 584 | GlnAspPrometGlnLeuArgProGlyProPro | ----- | PheProGlyProArg | | 599 |
| QY | 4104 | ACAACCGCAGCAGTCATGGGATGCGGACCACTCT | ----- | CCTAGCATGGAT | | 4151 |
| Db | 600 | ----- | PheProGlyAsnGlnMetGlnArgValProGlyPheGly | | | 612 |
| QY | 4152 | GGAACAGATCATTTATCTGCATCTGTTCCG | --- | CAAGCTAATACTTCGACGGTTCAGGCA | | 4208 |
| Db | 613 | GlyMetGlnSerMetPrometGluValPrometAsnAlaMetGlnArgProValArgPro | | | | 632 |
| QY | 4209 | GGCACAACA | ----- | ACAGTGCCTCTCAGCAACAAGAACTGTTTCAGGCA | | 4253 |
| Db | 633 | GlyMetAlaTrpAsnGluAspLeuProProIleGlyGlyProSerAsnPheAlaGlnAsn | | | | 652 |
| QY | 4254 | GACACCCCATCGCCGTCAAATCAAAATCGTAGTAGAAATACCGGATCGTCAAGCGTT | --- | | | 4310 |
| Db | 653 | AlaValProTyrrProGlyGlyGlnGlyGluAlaGluArgPheMetThrProArgValArg | | | | 672 |
| QY | 4311 | ----- | CTTACGCATACTTAAGCAGCAACCAAGTACCCCTTATCTCATCTATCC | | | 4361 |
| Db | 673 | GluGluLeuLeuArgHisGlnLeuLeuGluLysArgSerMetGlyMetGln | ----- | | | 689 |
| QY | 4362 | CCAAAGGAATTTGAGTCTTTCCGTCAGTCCCTCTGCTGTATGTATATTTGTTAAATTT | | | | 4421 |

| | | | |
|----|------|---|------|
| Db | 690 | -----ArgProLeuGlyMetAlaGlySerGlyMet----- | 699 |
| QY | 4422 | TTTAAGACAAATCAATAATGAATTGCGTTAATAATAGTTATATATTACATAACTCGGA | 4481 |
| Db | 700 | -----GlyGlnSerMetGlu-----MetGlu | 706 |
| QY | 4482 | AATTGTATGAAAAAATCAGAAATAGAAAAAATAAATTATTTCGGA----- | 4529 |
| Db | 707 | ArgMetIleGlnAlaHisArgGlnMetAspProAlaMetPheProGlyGlnMetThrGly | 726 |
| QY | 4529 | ----- | 4529 |
| Db | 727 | GlyAspGlyLeuAlaGlyThrProMetGlyIleGluPheGlyGlyArgGlyLeuLeu | 746 |
| QY | 4530 | ---CCGCC-----ATCCATTCTTGAAATCCAAATTTCTGGA----- | 4562 |
| Db | 747 | SerProPheMetGlyGlnSerGlyLeuArgGluValAspProPheMetGlyProGlyAsn | 766 |
| QY | 4563 | -----GTGATGTGTAGAGATATCTACTATTAAATTAACAAGAAATTCATATCCGT | 4616 |
| Db | 767 | LeuAsnMetAsnMetAsnValAsnMetAsnMetAsnMetAsnLeuAsnValGlnMet-Th | 786 |
| QY | 4617 | TAATGAAAAATCACTATTGTGTTTAATAAGAAATTAATAATATGTTATTATATATTCTA | 4676 |
| Db | 786 | rProGlnGlnGlnMetLeuMetSerClnLysMet-Arg-----GlyP | 800 |
| QY | 4677 | CAGGTGATTAACATGAAAAAGTAGCGCAAGCCACAGGGTCAGCGTCACAGTAATA | 4736 |
| Db | 800 | roGlyAspMetMet-----GlyProGlnGly--LeuSerPro----- | 811 |
| QY | 4737 | GTCTAATAGAGCAATAATAAGATGTCATTTGCTGCATCCAGTCTGTTTAACCCGC | 4796 |
| Db | 812 | -----GluGluMetAlaArgValArg--AlaGlnAsnSerSerGlyMetMetGlyG | 828 |
| QY | 4797 | ATCCACATATG-----CAAGCAATTCAAAAT | 4823 |
| Db | 828 | lyProGlnLysMetLeuMetProSerGlnPheProAsnGlnGlyGlnGlyPheSerG | 848 |
| QY | 4824 | CAGCATTAACGCTTATAA-----ATGGGCTCTACCAATATACAGATGG | 4868 |
| Db | 848 | lyGlyGlnGlyProTyrGlnAlaMetProGlnAspMetGlyAsnThrPro----- | 864 |
| QY | 4869 | AGGTAATATTTAAATATTTTATTTAACGTTTGTGTATTTACTTCTTTTCAGCG | 4928 |
| Db | 865 | -----AspMetPheSerP | 869 |
| QY | 4929 | T---CAAGCATCAGCGCAAGGTGATCCGTACAATTTAGTCGGCGCTCCGATAATATTC | 4984 |
| Db | 869 | roAspGlnSerSerValProMetGlyThrValGlyThrAlaArgLeuSer--HisMetP | 888 |
| QY | 4985 | CGCTAAATCCCAATAGTGGCAATCGGCGCCACCAACAAGATGACCCAAAACTTTCGATC | 5044 |
| Db | 888 | roLeuProProAlaSer--AsnProProGlySerValHisLeuAlaSerAsn----- | 904 |
| QY | 5045 | CAATCTCTTCTTGGCACAAATGTCCCAACTAACAAGTTGCGTGTCCAGCATGGGTA | 5104 |
| Db | 905 | ----ArgGlyLeuGlyArgArgProSerAspLeuThrIleSerIleAsnGlnMetGlys | 923 |
| QY | 5105 | GTCCAGCCGGAAGTGGTGTATGACGATGAGGGGTCGGGACCGTCGACATCAATA | 5164 |
| Db | 923 | erPro-----GlyMetGlyHisLeuLysSerProThrLeuSerGlnValHis- | 938 |
| QY | 5165 | TTGAGCATGAATAATTTCCGAGCATAGATGATCAGAAATAGATACATAAATCAAAATA | 5224 |
| Db | 939 | -----SerProLeuValThrSerProSerAlaAsnLeuLysSerProG | 953 |
| QY | 5225 | ACTGTCAATTCATGAATGTCGTATGAATCAATGGGTCC----- | 5265 |
| Db | 953 | lnThrProSerGlnMetValProLeuProSerAlaAsnProProGlyProLeuLysSerP | 973 |
| QY | 5266 | -----CGAATGC | 5272 |
| Db | 973 | roGlnValLeuSerSerSerLeuGlyValArgSerProThrGlySerProSerAlaGlyLeuL | 993 |

QY 5273 TGAATCCTAAATGTGCGTAGCAGCGGTCCAAATGACCGCCTGGCTT-----AATC 5326
Db 993 ySerProSerMetAlaValProSer-----ProGlyTrpValAlaSerP 1008
QY 5327 CTAATTC-----CCCAATGTGATTAAGAGAAAT-----TCCATAG 5365
Db 1008 roLySerThrAlaMetProSerProGlyValSerGlnAsnLysGlnProProLeuSerIleA 1028
QY 5366 GGTCTGGCTGTGGCTCAGCAAACTTTCAAGGGGTTGTTCCACCTGTGTGC- 5424
Db 1028 snSerSerSerThrLeuGlyAsnValGlu-----GlnGlyAlaLeuProProSerAlaP 1046
QY 5425 -----AGAATGATGGGTGCGAATGCCAGTCAATTTGGTTCGAAT---TTCATCCGAATA 5476
Db 1046 roArgAsnSerSerSerAlaProProAlaAsnProSerSerGlyLeuMetAsnPro---- 1064
QY 5477 TTCAGGTAAGGCGAGTACCCCAACACCATACATACATGCGCAGTAAGGCGACAGAAACG 5536
Db 1065 -----SerLeuProPheThrSerSerProAspProThrProSerGlnAsn- 1079
QY 5537 CCAACAACAATAACAACATGAGCTAATAATGTGCGAATGCCACCTAGTCTGGAATTTT 5596
Db 1080 -----ProLeuSerLeuMetMetS 1086
QY 5597 TGCAG-----AGTACGCTTAACCTCAAAATGGGTGCTGTAGCAATGGGTG- 5643
Db 1086 erGlnMetSerIleTyAlaMetProSerSerThrProLeuTyHisAsnAlaIleLysT 1106
QY 5644 -----CCAATATGCCCAACCATCAGCCA 5665
Db 1106 hrIleAlaThrSerAspAspGluLeuLeuProAspArgProLeuLeuProProPro- 1125
QY 5666 GCGACGGTACTCTCGGAATGCCAGGATTGATGCGCGGACCGAGAGCCGAGATATGCTAA 5725
Db 1126 -----ProProGlnGlySerGlyProGlyIleSerAsn----- 1136
QY 5726 TGAATTCTTCCGAGAGCAACACAGAACAGATCACAACAATCCTGGGCGCAAGCAAT- 5784
Db 1137 --AsnGlnProAsnGlnMetHisMet-----AsnProAlaAlaIaGlnS 1151
QY 5785 -----GGTATTAACTTCTTTCAGAAATGCAATCAATGTCTATTGTTGACGAAGAG 5836
Db 1151 erProMetGlyMetAsn----- 1156
QY 5837 GTGATTACCGCGCATGACGCGATCAATGATATTTGTCACCATCTATGATAAGGGCA 5896
Db 1157 -----LeuProGlyGlnGlnProLeuSerHisGluProProProThrMetLeuProSerP 1175
QY 5897 TGGCTCCACATGCC-----ATCGCGCCAAATGTAATGGGTGCGCGGATGCCAC 5944
Db 1175 roThrProLeuGlySerAsnIleProLeuHisProAsnAlaGlnGlyThrGlyGlySers 1195
QY 5945 CCGTTAACAGGCAAAATTCAGTTTGACAGATCATCGGATGTATTGACTGTGTCCGGGATC 6004
Db 1195 erGlnAsnSerMetMetMet-----AlaProGlyGlyP 1206
QY 6005 CGTCATCATTTTTCACCTAACGCTTCTCTGCAACAGCGCTGGACACACATGTTTGATCAG 6064
Db 1206 roAspSerLeu-----AsnAlaProCys-----GlyProValProSerSerSerg 1221
QY 6065 CACAACAGGCCAATGACGCTTAAGACACAACAC----- 6096
Db 1221 lInMetMetSerPheProProArgLeuGlnGlnProHisGlyAlaMetAlaProThrGlyA 1241
QY 6097 -----ATAAAGACATACCTAGTGAATGTGTCAAAACCAATCCGGGAC 6139
Db 1241 laGlyGlyProGlyLeuGlnGlnHisTyProSerGlyMetAlaLeuProProGluAspL 1261
QY 6140 TTGCAGTGGCACAAGGCGAGATCCAACTGCATGGCGAAGGACATGCGCAGGGTCAGTCTT 6199
Db 1261 euPro---ThrGlnProProGlyProIleProProGlnGlnHisLeuMetCylLylVgLYM 1280

QY 6200 TAAATGGACCTACTAATAATAATTAATGTCACTGCGGAGAGTGTCACTACTAAGC 6259
Db 1280 etThrGly-----ArgMetGlyAspAlaIleTyProProGlyValLeuProG 1295
QY 6260 GTGTCTCTGGCATC-----AATTCGTAGGTCCCTCTTCTAGC- 6297
Db 1295 lYValAlaSerValLeuAsnAspProGluLeuSerGluValIleArgProThrProThrG 1315
QY 6298 -----GACCTGAAGTAT-----GCCACGCAATATCATAGTTTTCAGC 6334
Db 1315 lYIleProGluPheAspLeuSerArgIleIleProSerGlyLysProSerSerThrLeuG 1335
QY 6335 AGCAGTTATATGCTACCAACACAGAGTCAA-----CAACAACAGCATATATGC 6382
Db 1335 lIn-----TyrrPheProLysSerGluAsnGlnProProLysAlaGlnProProAsnLeuH 1353
QY 6383 ACCAGCAGCACCGACGACCATGATA-----ACAATGCCCGCGGAATTTAT 6427
Db 1353 lSLeuMetAsnLeuGlnAsnMetMetAlaGluGlnThrProSerArgProProAsnLeu- 1372
QY 6428 CACCAATCCAAACGTTCTTTGTGCAACAATAACTTCTAAATTTTGGCCGCCCT 6481
Db 1373 --ProGly-GlnGlnGlyValGlnArgGlyLeuSerMetSerMetCysHisPro 1389

RESULT 14

AAB36095
ID AAB36095 standard; protein, 1596 AA.
XX
AC AAB36095;
XX
DT 26-JUN-2003 (first entry)
XX
DE Drosophila sp. Mastermind (Mam) protein.
XX
KW Notch signal transduction; Mastermind; Mam; Mip1; Mip30; Mip6; cancer;
KW gene therapy; fruitfly; cyostatic.
XX
OS Drosophila sp.
XX
PN WO2002102987-A2.
XX
PD 27-DEC-2002.
XX
PF 18-JUN-2002; 2002WO-US019189.
XX
PR 18-JUN-2001; 2001US-0299112P.
XX
PA (GENO) GEN HOSPITAL CORP.
XX
PI Artavanis-Tsakonas S, lake RJ;
XX
DR WPI; 2003-167509/16.
DR N-PSDB; AAD54620.
XX
PT Inhibiting or agonizing Notch signal transduction in a cell for preparing
PT a composition for treating or preventing cancer comprises contacting the
PT cell with an antagonist or agonist of sumolation.
XX
PS Disclosure; Fig 1; 84pp; English.
XX
CC The invention relates to a method for modulating Notch signal
CC transduction in a cell which comprises contacting the cell with an
CC antagonist or agonist of sumolation. The invention is based on the
CC interactions of Mastermind (Mam) protein with the Mip1, Mip30 and Mip6
CC proteins. The method is useful for preparing a composition for treating
CC or preventing a disease e.g. cancer. The invention is also used in gene
CC therapy. The present sequence is Drosophila sp. Mam protein used to
CC illustrate the method of the invention
SQ Sequence 1596 AA;
Alignment Scores: 1.84e-13 length: 1596
Pred. No.:

| | | | |
|-------------------|--|--|--------------|
| QY | 6186 | GCAGGGTCAGTCTTTAAATTGACCTACTAATAATAATTAAATGTCAACTGCCGGAAGTGT | 6245 |
| | | | |
| Db | 1368 | oMetArgGlnGlyProGlyGlyPheAsnGlyGlyAsnDhewetProAsn----- | 1384 |
| QY | 6246 | CAGTCTACTAACCAGGTGTCTCTGGCATCAAT-----TTCTAGGTCC | 6287 |
| | | | |
| Db | 1385 | -GlyAlaProAsnGlyAlaAlaGlySerGlyProAsnAlaGlyGlyMetMetThrGlyPr | 1404 |
| QY | 6288 | CTCTTCTACGACCTGGAATGATGCCAGCAATATCATAGTTTTCAGCAGCAGTTATATGC | 6347 |
| | | | |
| Db | 1404 | oAsnValProGlnMetGlnLeuThrProAla-----GlnMetGlnGlnGlnHis----- | 1420 |
| QY | 6348 | TACCAACACCAGAAGTCAACACACACAGCATATGCACGACGACAGCAGCAACATGAT | 6407 |
| | | | |
| Db | 1421 | -----MetArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnHis----- | 1435 |
| QY | 6408 | AACAATGCCCGCCGAATTATCACCAAT | 6435 |
| | | | |
| Db | 1436 | ---MetGlyProGlyAlaAlaAsnAsn | 1443 |
| RESULT 15 | | | |
| ABB63688 | ID ABB63688 standard; protein; 1594 AA. | | |
| XX | AC | ABB63688; | |
| XX | DT | 26-MAR-2002 (first entry) | |
| XX | DE | Drosophila melanogaster polypeptide SEQ ID NO 17856. | |
| XX | KW | Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical. | |
| XX | OS | Drosophila melanogaster. | |
| XX | PN | WO200171042-A2. | |
| PD | 27-SEP-2001. | | |
| XX | PF | 23-MAR-2001; 2001WO-US009231. | |
| XX | PR | 23-MAR-2000; 2000US-0191637P. | |
| PR | 11-JUL-2000; 2000US-00614150. | | |
| XX | PA | (PEKE) PE CORP NY. | |
| XX | PI | Venter JC, Adams M, Li PWD, Myers EW; | |
| XX | DR | WPI; 2001-656860/75. | |
| DR | N-PSDB; ABL07791. | | |
| PT | New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signaling and cell-cell interactions. | | |
| XX | PT | | |
| XX | PT | | |
| PS | Disclosure; SEQ ID NO 17856; 21pp + Sequence Listing; English. | | |
| XX | | | |
| CC | The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signaling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB2072). The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences | | |
| CC | | | |
| CC | | | |
| XX | SQ Sequence 1594 AA; | | |
| Alignment Scores: | | | |
| Pred. No.: | | 2.24e-13 | Length: 1594 |
| Score: | | 269.50 | Matches: 295 |

[illegible]

Db 493 AsnSerGluAsnAsnAspThrPheLeuAspLeuIleAsnAsnLeuGlnAspPheAsnPro 512
QY 3474 ---CATATGCCAGGTAACGCCAAATCGAGCTTATCG---GCCACAAGTTCAGGACTT 3527
Db 513 SerPheLeuAspGlyPheAspGlyLeuProLeuAspIleLeuThrGluAspGlyIle 532
QY 3528 TCGGAAGATGTAATGCATCCAGGGGATGTTATATCAGATATGGTCCGTATAGATGT 3587
Db 533 LysValGluProProAsnAlaGlnAspLeuIleAsnSerLeu----- 546
QY 3588 AATAATAATCAAAAAACAGTGTGCAATGTGAGTCTGAGTAGGTGTGTC----- 3638
Db 547 -----AsnValLysSerGluGlyGlyLeuGlnHisGlyPheGlyGlyPheGlyLeuGly 564
QY 3639 -----ACTGAACAACATGCAGCTGAGTAATGTCAATATGCAATGCTCAAGCTCC 3689
Db 565 LeuAspAsnProGlyMetLysMetArgGlyGlyAsnProGlyAsnGlnGlyPhePro 584
QY 3690 GGGCCCCGCAATGCCAATATGATGGGAAGCTTAACGATATGCTAGCTCGTTGGCAAC 3749
Db 585 AsnGlyProAsnGlyGlyThrGlyGlyAlaProAsn-----AlaGlyGlyAsnGlyGly 602
QY 3750 ACAAGCTGCAACGTTCATCGGAACGGCCCCAGATATGTTAAAGAACTTTAAATCAAGAT 3809
Db 603 AsnSerGlyAsnLeuMetSerGlnHisPro---LeuAlaAlaGlnThrLeuLysGlnMet 621
QY 3810 AGCCGAACCCATTCATCATCAAGGGGAGTTGCTCAATGAGTGGTGAAGATTCAACAT 3869
Db 622 AlaGluGlnHisGlnHisLysAsnAlaMetGlyGlyMet----- 634
QY 3870 CAATTTTTCGAAGAAGCCTCAAGGGGGGCAAGCCGACAAGTCACTGAACTGAGTA 3929
Db 635 -----GlyGlyPheProArgProProHisGly---MetAsn 645
QY 3930 CCACAACAGCAAAACCCCTTGTGATGTGGTGAACCTCGTTAAACAACAGGTGCGACCC 3989
Db 646 ProGlnGlnGlnGln-----GlnGlnGlnGlnGlnGln 656
QY 3990 CTGCAAGGTCCACCTCTCTTACCACCTCCATCCAGAGATCTGCGTCAATCAATAGCC 4049
Db 657 GlnGlnGlnGlnAlaGlnGlnGlnHisGlyGlnMetMetGlyGlnGlnProGlyArg 676
QY 4050 ACTCAATGCCCAATCCCTCGAGTCCAAACAATCTATCTCTC-----CCGTCAACCGCG 4103
Db 677 TyrAsnAspTyrGlyGlyGlyPheProAsnAspPheGlyLeuGlyProAsnGlyProGln 696
QY 4104 ACAACCCGACAGTCAATGAGATTGCCGACCAACTCTCTAGCATGATGGAACAGATCA 4163
Db 697 -GlnGlnGlnGlnAlaGlnGlnGlnPro-----GlnGlnGlnHis 710
QY 4164 T----- 4164
Db 710 sLeuProProGlnPheHisGlnGlnLysGlyProGlyAlaGlyMetAsnValGly 730
QY 4165 -----TATCTGATCTGTTCGCAAGCTATA-----CT 4193
Db 730 nglnAsnPheLeuAspIleLeuGlnGlnLeuPheTyrSerSerGlnAsnAspPheAspLe 750
QY 4194 TCGACGGTTCAGGCAAGCAACAACAACAGTCTCTCAGCAACAACAAGACTTTTTCAGGCA 4253
Db 750 uLysArgLeuGlnGlnGlnAlaMetGlnGlnGln-----GlnGln 765
QY 4254 GACACCCCATCGCCGTCAAATCAAAATCGTAGAGAAATACCGGATCGTCAAGCGTTCTT 4313
Db 765 nglnHisHisGlnGlnGln-----GlnGln 773
QY 4314 ACGCATACTTAAGCAACAACCAAGTACCCCTTATCTCATCTATCCCAAGAATTT 4373
Db 774 -----GlnGlnGlnProLys-----MetGlyGlyValPro----- 783
QY 4374 GAGTCTTTCGGTCACTCTCTGCTGTATGTATATTGTTAATTTTAAAGACAAA 4433
Db 783 ----- 783

QY 4434 TCAATATGAATTGCGTTAATGATAAGTTATATATATACATCACTCGGAATTTGATAGAA 4493
Db 784 ----- 788
QY 4494 AAAATCAGGAATAGAAAAATAATATTATTTCCGGACCGCCCATTCATTCTTGAATCCA 4553
Db 788 lnglnGlnGln----- 791
QY 4554 ATTTCTGAGTGAATTGTTAGAGATAATCTACTATTAAATTAAACAGAAATTTCATATC 4613
Db 791 ----- 791
QY 4614 CGTTAATTGAAATCAGTATTGTTTAAAGAAATTAAATAATGTTTATTATATATTT 4673
Db 791 ----- 791
QY 4674 CTACAGGTGATAACATGAAAGTAGCGCAACCAAGCCACAGGTCAAGGTCAACAGTAA 4733
Db 792 -----GlnGlnValProGlnGlnGlnLeuGlnGlnGln 803
QY 4734 ATAGTCTAATAGAGCAATAAAGATGTACGATTTGCTGCATCCAGTCCGTGTTTAACC 4793
Db 803 lnglnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 823
QY 4794 CGCATCCACATATGCAAGCAATTCAAATTCAGCATTAACGCCCTATAAATGGCTCTA 4853
Db 823 lnglnProAsnAlaAlaAlaAsn-----PheLeuAsnGlyProProArgGlyGlyP 840
QY 4854 CCAATAT---ACAGATGAGGTAAATATTAAATATTTTAATATTTAACGTTTGTGTTAA 4909
Db 840 roAsnGlyAsnGlnGlnProGlyAsn----- 848
QY 4910 TTTATCTTCTTTTTCAGCGTCAAGCATCAGCCGAAGGTGATCCGTAATTTAGTCGC 4969
Db 849 -----LeuAlaGlnGlnGlnGlnGlnProGlyAlaGlyProGlnGlnGlnGlnAla 866
QY 4970 GC-----TCCGATAATATTCGCTAAATCCCAATAGT-----GCCAATCGGC 5011
Db 866 rgGlyAsnAlaAlaAsnGlyGlnGlnAsnAsnProAsnAlaGlyProGlyGlyAsnThrP 886
QY 5012 CGCCACCAACAAGATGACCCCAAACTTCGATCCAAATCTCTTGTGGCACAATGTCCC 5071
Db 886 roAsnAlaProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 905
QY 5072 AACAACTAACAAAGTTGCGTGTCCAGCATGGGTAGTCCAGCCGGAAGTGGTATGACGA 5131
Db 905 lngThrGlnGlnLeuHisLysSerGlnGlnGly-----GlyGlyAlaHisGlyIleGln 923
QY 5132 TGATGGGGGTCCGGGACCGTCCGACATCAATATTGAGCATGGAATAATTCCGGAGCTAG 5191
Db 923 alSerAlaGly-----GlnHisLeuHisLeuSerGlyAspM 935
QY 5192 ATGATCAGGAATGATATACCAATAATCAAAATAACTGTCAATGATGATGATGATGATGA 5251
Db 935 etLysSerAsnValSerValAlaAlaGlnGlnGly-----ValPheP 949
QY 5252 ACTCAATGGGTCCCGAATGCTGAATCTAAATGTGCGTAGCAGCGCGTCCAATGAC 5311
Db 949 heserGlnGlnGlnAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 969
QY 5312 CG---CCTGGCTTAACTCTAATTCGCCCAATGGTGAATTAAGAGAAATTCATAGCGT 5368
Db 969 roAsnProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 989
QY 5369 CTGGCTGTGGC----- 5379
Db 989 alGlyValGlyValGlyValGlyValGlyValGlyValGlyValGlyValGlyValGlyVal 1009
QY 5380 --TCAGCAAACCTTCAAACTTTCAAGGGGTGTTCCACCCTGGTGCAGAAATGATGGGTC 5437
Db 1009 roAsnGlnAsnMetSerAsn---AlaAsnValProSerAspGlyPheSerLeuSerGlnS 1028

QY 5438 GAATGCCAGTCAATTTTGGT-----TCGAATT 5464
Db 1028 ergInserMetAsnPhaSnGlnGlnGlnGlnAlaAlaGlnGlnGln 1048
QY 5465 TCAATCCGAATATT-----CAGTAAAGCGAGTACCCCAACACCATCAATACA 5515
Db 1048 alGlnProAsnMetArgGlnArgGlnThrGlnAlaGlnAlaAlaAlaAla 1068
QY 5516 TGCCAGTAAGGCGACAGAACGCCAACACAATACACAATGAGCTAATAATGTCCGAA 5575
Db 1068 laAlaAlaGlnAlaGlnAlaAlaAlaAsnAlaSerGly----- 1080
QY 5576 TGCCACCTAGTCTGGAATTTTGCAGAGGTACGCTAACCCCTCAATGGGT----- 5625
Db 1081 -----ProAsnValProLeuMetGlnGln-----ProGlnValGlyValG 1096
QY 5626 -----GCTGTAGGCAATGGGTGCCCAATATATGCCCATCAG 5662
Db 1096 lyValGlyValGlyValGlyValGlyValGlyValGlyValGlyProGlys 1116
QY 5663 CCAGCGACGGTACTCTCGAATGCCAGATGATGCGGACCA-----GAGCGCGAG 5716
Db 1116 ergGlyGlyProAsnAsnGlyAlaMetAsnGlnMetGlyGlyProMetGlyGlyMetProG 1136
QY 5717 GTATGCTAATGAATTTCTCCGAGAGCAACACCAAGATCACAACAATCCTGGGG 5776
Db 1136 lyMetGlnMetGlyGlyPro-----MetAsnProMetGlnMetAsnProAsnA 1152
QY 5777 CAAGCAATGGTATTACTTTCTTCAGAATTGCAATCAATGCTATTGTGACGAAGAG 5836
Db 1152 laAlaGly-----ProThrAlaGlnGlnMetMetGlySerGlyAlaG 1167
QY 5837 GTGGA-----TTACCCCGGCCATGACGGATCAATGAAT-----A 5869
Db 1167 lyGlyProGlyGlnValProGlyProGlyGlnGlyProAsnProAsnGlnAlaLysPheL 1187
QY 5870 TTGGTCAACCATTAATGATAAGGGGCGATGCGTCACATGCCATGCGCCAAATGTAATGG 5929
Db 1187 euGlnGlnGlnGlnMetMetArgAlaGlnAlaMetGlnGlnGlnGlnHisMetSerg 1207
QY 5930 GTGCGCGATGCCAACC-----GTTAACAGGCAATTCAGTTTGACACAGT 5974
Db 1207 lyAlaArgProProProGluTyrAsnAlaThrLysAlaGlnLeuMetGlnAlaGlnm 1227
QY 5975 CATCGAT-----G 5983
Db 1227 etMetGlnGlnThrValGlyGlyGlyValGlyValGlyValGlyValG 1247
QY 5984 GTATTGACTGTGTCGGGATCCGTCA--TCATTTTCACTAACGCTTCCTGCAACACGG 6040
Db 1247 lyValGlyGlyValGlyGlyAlaAsnGlyGlyArgPheProAsnSerAlaAlaGlnAlaA 1267
QY 6041 CT-----GACCACACATG----- 6054
Db 1267 laAlaMetArgMetThrGlnGlnProIleProProSerGlyProMetMetArgProG 1287
QY 6054 ----- 6054
Db 1287 lnhIsAlaMetTyrMetGlnGlnHisGlyGlyAlaGlyGlyGlyProArgThrGlyMetG 1307
QY 6055 -----TTTGATCAGCACACAGGCCAATCAGC 6082
Db 1307 lyValProTyrGlyGlyGlyAlaGlyGlyProMetGlyGlyProGlnGlnGlnAlaArgP 1327
QY 6083 CTAAGACACACACATA-----AAGACATACCTAGTGAATGTGTCAAAACCAATCGG 6136
Db 1327 roProAsnValGlnValThrProAspGlyMetProMetGlySerGlnGlnGluTrrArgH 1347
QY 6137 GACTTGACGTGGCACAAGGCGACATCCAACGT-----CATGGGCAAGGACATG 6184
Db 1347 isMetMetMetThrGlnGlnGlnThrGlnMetGlyPheGlyGlyProGlyProGlyGlyP 1367
QY 6185 CGCAGGGTCAGTCTTTAATTGGACCTACTAATAATAATTAAATG----- 6228

Db 1367 rometArgGlnGlyProGlyGlyPheAsnGlyGlyAsnPhemetProAsnGlyAlaProA 1387
QY 6229 --TCAACTGCCGGAAGTGTCAAGTGTCTACTTAACGCT--GTCTGTGGCATCAATTCGTAG 6283
Db 1387 snGlyAlaAlaGlySerGlyProAsnAlaGlyGlyMetMetSerGlyProAsnValProG 1407
QY 6284 GTCCCTCTTCTACGGAACCTGAAGTATGCCGCAATATCATAGTTTTCAGCAGCAATTAT 6343
Db 1407 lmetGlnLeuThrProAlaGlnMetGlnGlnGlnLeuMetArgGlnGlnGln---- 1425
QY 6344 ATGCTACCAACACAGAAGTCAACAACAACAACATATGCAACGACGACGACGACGACAA 6403
Db 1426 -----GlnGlnGlnGlnGlnGlnHisMetGlyProGlyAlaAlaAsnAsm 1442
QY 6404 TG 6405
Db 1442 et 1442

Search completed: November 3, 2005, 01:02:05
Job time : 1116.5 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: November 2, 2005, 22:24:08 ; Search time 1054 Seconds
(without alignments)
6713.387 Million cell updates/sec

Title: US-10-664-859-1
Perfect score: 12037
Sequence: 1 acgagtcgtctcttattat.....aaaaaaaaaaaaaa 6909

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1612378 segs, 512079187 residues

Total number of hits satisfying chosen parameters: 3224756

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p.model -DEV=xlp
-Q=/cgn2.1/USPTO_spool_p/US1064859/runat_02112005.171807.3627/app_query.fasta_1.7047
-DB=UniProt -QFMT=fastan -SUFFIX=rup -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOCCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pcio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US1064859_@CGN_1_1295_@runat_02112005.171807.3627 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : UniProt_03:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|--------------|--------------------|
| 1 | 7280 | 60.5 | 1469 | 1 BCL9_DROME | O961d9 drosophila |
| 2 | 338.5 | 2.8 | 1530 | 2 Q67FY3 | O67fy3 brachydanio |
| 3 | 322.5 | 2.7 | 1723 | 2 Q8T248 | O8t248 dictyosteli |
| 4 | 320 | 2.7 | 1426 | 1 BCL9_HUMAN | O00512 homo sapien |
| 5 | 313 | 2.6 | 1474 | 2 Q67FY0 | O67fy0 brachydanio |
| 6 | 301 | 2.5 | 2232 | 2 Q81FX6 | O81fx6 caenorhadi |
| 7 | 297.5 | 2.5 | 1457 | 2 Q641L9 | O641l9 mus musculu |
| 8 | 296.5 | 2.5 | 1494 | 2 Q67FY2 | O67fy2 mus musculu |
| 9 | 296.5 | 2.5 | 1494 | 2 Q617B5 | O617b5 mus musculu |
| 10 | 289.5 | 2.4 | 1425 | 2 Q67FX9 | O67fx9 mus musculu |
| 11 | 288 | 2.4 | 4662 | 2 Q81LR9 | O81lr9 plasmodium |
| 12 | 281 | 2.3 | 1494 | 2 Q67FY1 | O67fy1 homo sapien |
| 13 | 280 | 2.3 | 1499 | 2 Q86U00 | O86u00 homo sapien |
| 14 | 279 | 2.3 | 1774 | 2 Q81AU8 | O81au8 plasmodium |
| 15 | 277.5 | 2.3 | 1902 | 2 Q640Q1 | O640q1 mus musculu |
| 16 | 277.5 | 2.3 | 3006 | 2 Q7YY56 | O7yy56 cryptospori |

| | | | | | |
|----|-------|-----|------|--------------|--------------------|
| 17 | 274.5 | 2.3 | 1918 | 2 Q86AF5 | O86af5 dictyosteli |
| 18 | 270.5 | 2.2 | 1596 | 1 MAM_DROME | P21519 drosophila |
| 19 | 269.5 | 2.2 | 1594 | 2 Q9V6W7 | O9v6w7 drosophila |
| 20 | 267.5 | 2.2 | 1902 | 2 Q925Q1 | O925q1 mus musculu |
| 21 | 261.5 | 2.2 | 2206 | 2 Q7QAV8 | O7qav8 anopheles 9 |
| 22 | 261.5 | 2.2 | 2235 | 2 Q81B09 | O81b09 plasmodium |
| 23 | 258.5 | 2.1 | 1655 | 2 Q24754 | O24754 drosophila |
| 24 | 258.5 | 2.1 | 2151 | 2 Q9VPL6 | O9vp16 drosophila |
| 25 | 256.5 | 2.1 | 5322 | 2 Q9VPL9 | O9vp19 drosophila |
| 26 | 255.5 | 2.1 | 2151 | 2 Q9Y1L3 | O9y1l3 drosophila |
| 27 | 254.5 | 2.1 | 1348 | 2 Q86KE2 | O86ke2 dictyosteli |
| 28 | 254.5 | 2.1 | 2038 | 2 Q9W3L3 | O9w3l3 drosophila |
| 29 | 251.5 | 2.1 | 1502 | 2 Q81S10 | O81s10 dictyosteli |
| 30 | 249.5 | 2.1 | 2038 | 1 FSH_DROME | P13709 drosophila |
| 31 | 249 | 2.1 | 1685 | 2 Q7REY7 | O7rey7 plasmodium |
| 32 | 248 | 2.1 | 3443 | 2 Q8JZM8 | O8jzm8 mus musculu |
| 33 | 247.5 | 2.1 | 1902 | 1 SMF1_HUMAN | O14497 homo sapien |
| 34 | 247.5 | 2.1 | 1999 | 2 Q8NFD6 | O8nfd6 homo sapien |
| 35 | 247 | 2.1 | 1255 | 2 Q9W3Q5 | O9w3q5 drosophila |
| 36 | 246.5 | 2.0 | 2634 | 2 Q63KC6 | O63kc6 burkholderi |
| 37 | 244.5 | 2.0 | 2072 | 2 Q7RSW8 | O7rsw8 plasmodium |
| 38 | 243.5 | 2.0 | 1208 | 2 Q96TA9 | O96ta9 homo sapien |
| 39 | 243.5 | 2.0 | 1323 | 2 Q7KSZ0 | O7ksz0 drosophila |
| 40 | 243.5 | 2.0 | 1374 | 2 Q7YU77 | O7yu77 drosophila |
| 41 | 243.5 | 2.0 | 1376 | 2 Q9VI63 | O9vi63 drosophila |
| 42 | 242 | 2.0 | 1353 | 2 Q7RDU8 | O7rdj8 plasmodium |
| 43 | 241 | 2.0 | 796 | 2 Q6NRE2 | O6nre2 xenopus lae |
| 44 | 241 | 2.0 | 1606 | 2 Q63UB0 | O63ub0 burkholderi |
| 45 | 241 | 2.0 | 2280 | 2 Q9V8E6 | O9v8e6 drosophila |

ALIGNMENTS

| RESULT 1 | BCL9_DROME | STANDARD; | PRT; 1469 AA. |
|----------|--|-----------|---------------|
| ID | BCL9_DROME | | |
| AC | O961D9; Q9V4D2; | | |
| DT | 28-FEB-2003 (Rel. 41, Created) | | |
| DT | 28-FEB-2003 (Rel. 41, Last sequence update) | | |
| DT | 25-JAN-2005 (Rel. 46, Last annotation update) | | |
| DE | Bcl-9 homolog (Legless protein). | | |
| GN | Name=lg9; Synonyms=BCL9; ORFNames=CG2041; | | |
| OS | Drosophila melanogaster (Fruit fly). | | |
| OC | Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; | | |
| OC | Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; | | |
| OC | Ephydroidea; Drosophilidae; Drosophila. | | |
| OX | NCBI_TaxID=7227; | | |
| RN | [1] | | |
| RP | SEQUENCE FROM N.A. | | |
| RC | STRAIN=Berkeley; | | |
| RX | MPB41NE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185; | | |
| RA | Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D., | | |
| RA | Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F., | | |
| RA | George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N., | | |
| RA | Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X., | | |
| RA | Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D., | | |
| RA | Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G., | | |
| RA | Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D., | | |
| RA | Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M., | | |
| RA | Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S., | | |
| RA | Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P., | | |
| RA | Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I., | | |
| RA | Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., | | |
| RA | de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M., | | |
| RA | Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., | | |
| RA | Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W., | | |
| RA | Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K., | | |
| RA | Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., | | |
| RA | Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J., | | |
| RA | Hoeltin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C., | | |
| RA | Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A., | | |
| RA | Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z., | | |
| RA | Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X., | | |

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mlehnina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Piltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svrtkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP GENOME REANNOTATION.
RX MEDLINE=22426069; PubMed=12537572; Matthews B.B., Campbell K.S.,
RA Misra S., Crosby M.A., Mungall C.J., Mungall K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celisner S.B., de Grey A.D.N.J., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review.";
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley; Tissue=Embryo;
RX MEDLINE=22426066; PubMed=12537569;
RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA George R.A., Guarin H., Kronmiller B., Pacleb J.M., Park S., Wan K.H.,
RA Rubin G.M., Celisner S.E.;
RT "A Drosophila full-length cDNA resource.";
RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
RN [4]
RP SEQUENCE OF 6-1469 FROM N.A., AND MUTAGENESIS OF GLY-514; LEU-534 AND
RP ILE-537.
RX MEDLINE=21952490; PubMed=11955446; DOI=10.1016/S0092-8674(02)00679-7;
RA Kramps T., Peter O., Brunner E., Nellen D., Froesch B., Chatterjee S.,
RA Murone M., Zuelig S., Basler K.;
RT "Wnt/wingless signaling requires BCL9/legless-mediated recruitment of
RT pygopus to the nuclear beta-catenin-TCF complex.";
RL Cell 109:47-60(2002).
CC -1- FUNCTION: Involved in signal transduction through the Wnt pathway.
CC -1- SUBUNIT: Binds to ARM and PYGO.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- DEVELOPMENTAL STAGE: Expressed both maternally and zygotically
CC throughout development.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE003844; AAF59345.2; -;
DR EMBL: AY051651; AAK93075.1; -;
DR EMBL: AF457205; AAL91368.1; -;
DR FLYBase: FBgn0039907; 198.
DR GO: GO:0005634; C:nucleus; IDA.
DR GO: GO:0030528; F:transcription regulator activity; IPI.
DR GO: GO:0030177; P:positive regulation of Wnt receptor signal. . .; IPI.
DR GO: GO:0007367; P:segment polarity determination; IMP.
KW Developmental protein; Nuclear protein; Segmentation polarity protein;
KW Wnt signaling pathway.
FT DOMAIN 511 555 ARM-binding.
FT DOMAIN 1134 1173 Asn-rich.

FT DOMAIN 1340 1449 Gln-rich.
FT DOMAIN 1162 1169 Poly-Asn.
FT MUTAGEN 514 514 G->E: in allele 1gs-21L.
FT MUTAGEN 534 534 L->F: in allele 1gs-17E; segment polarity
FT phenotype.
FT MUTAGEN 537 537 I->K: in allele 1gs-17P.
SQ SEQUENCE 1469 AA; 153759 MW; 5672E01B7200ED08 CRC64;

Alignment Scores:
Pred. No.: 0 Length: 1469
Score: 7280.00 Matches: 1465
Percent Similarity: 73.08% Conservative: 1
Best Local Similarity: 73.03% Mismatches: 3
Query Match: 60.48% Indels: 537
DB: 1 Gaps: 5

US-10-664-859-1 (1-6909) x BCL9_DROME (1-1469)

QY 453 ATGCTCTCGACACATATGCCCGGAGTCCACCAACAAGCCGCAACCAACTCCGAT 512
DB 1 MetLeuSerThrThreMetProArgSerProThrGlnGlnProGlnProAsnSerAsp 20

QY 513 GCCTCTCAACAAGTGCATCTGATCAATCTCGAGCAGCCATCGGAATGGGACTCG 572
DB 21 AlaSerSerThrSerAlaSerGlySerAsnProGlyAlaAlaIleGlyAsnGlyAspSer 40

QY 573 GCGGCGACGAAGTCTCTCCGAAGACCTTATAGCGAACCTTTTCTACTTGTGCGCC 632
DB 41 AlaAlaSerArgSerSerProGlyThrLeuAsnSerGluProPheSerThrLeu-Ser- 59

QY 633 GGTAAGACTTGATTGATTCTCTTGTCCGAATTATAACAATTCTGTGTTCCAGA 692
DB 60 -----ProAs 61

QY 693 TCAATAAATTGACGCCAGAGGACTGAGAAAAGCGGACTATCACTAGTGATGA 752
DB 61 pGlnIleGlySerThrProGluGluGlyThrGluGlySerGlyLeuSerThrSerAsp 81

QY 753 AGCTGCCACTGAGAGAGCCCGAGGAGTGAATATCTGCCGAGGAGCAACTATGCT 812
DB 81 sAlaAlaThrGlyGlyAlaProGlySerGlyAsnAsnLeuProGluGlyGlnThrMetLe 101

QY 813 AAGGACGAATCTACGAGCACAATCACTGCTGCTAGTCCCTTCCACAAAACCTCCAG 872
DB 101 uArgGlnAsnSerThrSerThrIleAsnSerCysLeuValAlaSerProGlnAsnSerSe 121

QY 873 TGAACACTGAATAGCAGCAATGTCTGCTACAGTGGGCTTACTCAGATGTAGATTG 932
DB 121 rGluHisSerAsnSerSerAsnValSerAlaThrValGlyLeuThrGlnMetValAspCy 141

QY 933 TGACGACGAATGAAGAAAACAATGTAGTGTGAAGAGCAGGAGAGCTGTAGACTGCG 992
DB 141 sAspGluGlnSerIleSerIleSerIleSerValIleSerValIleSerValIleSerVal 157

QY 993 CCTACAATGCTTAAATTTTAAATGTATTGGCGTTACACCTTGTTAATCATTAATT 1052
DB 157 ----- 157

QY 1053 GTTTTTTTTGTCTACTACTTACAATTTAGTTTAACTTGTAACCTGACTAAACTCG 1112
DB 157 ----- 157

QY 1113 CGAAGCTCGGATCAAAACAGACATTTTCTTGAACCGTAATTAACTCAATAAAATATTA 1172
DB 157 ----- 157

QY 1173 ATTCACTTGATGGAATGATCATATCATAGATGTACTCAAAACATCTCAAGAAAGACCTCAAA 1232
DB 157 ----- 157

QY 1233 TTGATCACTAATTAGTTTGAGAAAATGTGCTGACTTTAAGATATATTAAATTAA 1292
DB 157 ----- 157

QY 1293 AAATTGCTGAGTAATGATATATAGTCAATAATTTTGTAAACTGCTAAAGCA 1352
 Db 157 ----- 157
 QY 1353 TTTTGAATAGCCGTGCTACGACAGATGCTACTAGACCGGTGTAAGCTAATTTTATT 1412
 Db 157 ----- 157
 QY 1413 AAAAGTGTCTTAATATTCCATTAACCATTAATGTCCTTCAGAAATAAGTTCTAATAA 1472
 Db 158 -----GluIleSerSerAsnIly 163
 QY 1473 AGCAAAAGGTCAAGCAGCTGTGCGGCTGCGAAACAGGTTCTACATCCAGTTGACTGT 1532
 Db 163 sAlaIySgIyGlnAlaIaIaGlyGlyCySgIuThrGlySerThrSerSerLeuThrVa 183
 QY 1533 CAAGGAAGAACCCACCGATGTCTTAGGCGTTTGTAAATATGAAGAAAGAAAGAGA 1592
 Db 183 llySgIuGlnProThrAspValLeuGlySerLeuValAsnMetIlySgIuGlnIuArgI 203
 QY 1593 AAATCATTCGCCCAACGATGTCCCTGTGGTTTGTCAATTGGTAATGCAACAGACAA 1652
 Db 203 uAsnHisSerProThrMetSerProValGlyPheGlySerIleGlyAsnAlaGlnAspAs 223
 QY 1653 CTCCGCTACACCGGGTAAGTTTAAAGATCCATATTAAGCAAAATACAGAATTAATGT 1712
 Db 223 nSerAlaThr----- 226
 QY 1713 CAGTTACCAATTTTATTGTAGTCAAGAACTACTATAGCATATCTCGCTTTTAA 1772
 Db 226 ----- 226
 QY 1773 TTTTATTTTAATTAGAAATACGAATATTTCTAATTTGTAAATAAATTGATTAA 1832
 Db 226 ----- 226
 QY 1833 CTAGAATTAAAAACCTTTTGAATTAGGACATACCCTTCCAAAAATCAGTAATCATTTGG 1892
 Db 226 ----- 226
 QY 1893 AACGAGATGTGTCCCGAAGAGACTACTATAAAACCTTTGAGCTATCTGATACTGCA 1952
 Db 226 ----- 226
 QY 1953 CGCTACTAAAAATGATTAGTTAGAAATGCGGTGAATTTTGTAGAAATTTTCATTTT 2012
 Db 226 ----- 226
 QY 2013 AGAAGAAATGTATTTTATTAAACCCCTTCAAGCGAATACTATTGTTCTACGATA 2072
 Db 226 ----- 226
 QY 2073 TTTTGGAAAAACAATGCTTAAGTTGAAAGTCCCTATAAAACAGAATTCCACGGTTTCA 2132
 Db 226 ----- 226
 QY 2133 AATACTAACCAAGTTTGTATTATTGATTAAATGAGAAATTATCACACTTCAGTTA 2192
 Db 226 ----- 226
 QY 2193 AAATGTTAATTCGATTAAAGTCGACAATCACAGAGATTTCATTTTGGGTGTATAT 2252
 Db 226 ----- 226
 QY 2253 ATAGAAGTCGCTTCAACTCTTCTGCGCGCTTCAACCACTACGTGAGTTCCGCCCGCA 2312
 Db 226 ----- 226
 QY 2313 GTGATTTATATAGATTTAAGAGATTATTTTATGTTATTTTAAATAATAT 2372
 Db 226 ----- 226

QY 2373 CTATATTATTCATTTTACATAGTTAAATTTGAAGAAATTTCAACGACAGTACCAGGAA 2432
 Db 227 -----Pro--ValIySileGluArgIleSerAsnAspSerThrThrGlu 240
 QY 2433 AAAAAAGATCGCTCTTGACAATGAATTAAGCAGAAATGACATGAGCGTCAATCAG 2492
 Db 241 LysIySgIySerSerLeuThrMetAsnAsnAspGluMetSerMetGluGlyCySAsnGln 260
 QY 2493 TTGAATCCCGATTTTATCAATGAATCTTAAATAATCCTGCAATTTGAGCATATTAGTA 2552
 Db 261 LeuAsnProAspPheIleAsnGluSerLeuAsnAsnProAlaIleSerSerIleLeuVal 280
 QY 2553 AGCGAGTAGAACCAATACCCGGAATCGAGTTGAGCGGGGACGGGAAATTTATTTACT 2612
 Db 281 SerGlyValGlyProIleProGlyIleGlyValGlyAlaGlyThrGlyAsnLeuLeuThr 300
 QY 2613 GCCAACGCCAATGGAATCTCTCGGGTAGCAGTAATTGTTGATTACATGCAACAGCAA 2672
 Db 301 AlaAsnAlaAsnGlyIleSerSerGlySerSerAsnCysLeuAspTyrMetGlnGlnGln 320
 QY 2673 AATCACATATTCGTGTTTCAACTCAGCTGCGCAACAAAGGGGCGCAATCAGTTTAAAGC 2732
 Db 321 AsnHisIlePheValPheSerThrGlnLeuAlaAsnIySgIyAlaGluSerValLeuSer 340
 QY 2733 GGTCAATTTTCAAACTATTATTGCGTATCACTGCACTCAGCCTGTCTCAAAAAGCTTCTG 2792
 Db 341 GlyGlnPheGlnThrIleIleAlaTyrHisCysThrGlnProAlaThrIySerPheLeu 360
 QY 2793 GAAGACTTTTTATGAAGAAACCCCTTAAAGATTAAACAAGTTACAGCGGACAAATCCGTC 2852
 Db 361 GluAspPhePheMetCysAsnProLeuIySileAsnIySleuGlnIuArgHisAsnSerVal 380
 QY 2853 GGTATGCCATGATAGGCATGGGGCAGGTGGAATACTCCTCTTAATCCTGTAGCCAAA 2912
 Db 381 GlyMetProIleIleGlyMetGlyGlnValGlyLeuThrProProAsnProValAlaIyS 400
 QY 2913 ATAAACAACAAGCAGCCACATACAAAGACGCTAGCGCTATTGAAACCCCAATTCATCA 2972
 Db 401 IleThrGlnGlnGlnProHisThrIySerThrValGlyLeuLeuIySProGlnPheAsnGln 420
 QY 2973 CATGAAACGACAAACGTAGTACTGTAGCGCGCTAGCAACTCTTTGTCAGACAGTCT 3032
 Db 421 HisGluAsnSerIySAspSerThrValSerAlaProSerAsnSerPheValAspGlnSer 440
 QY 3033 GATCCTATGGGCAACGAACTGAATGTGTCTGGGAAGCGGATCCTCAAAACACCACT 3092
 Db 441 AspPrometGlyAsnGluThrGluLeuMetCysTrpGluGlyGlySerSerAsnThrSer 460
 QY 3093 AGGTCTGGACAAACTCAGGAATCATGTAGACAGTATCATGATCCACGAGTCAACAG 3152
 Db 461 ArgSerGlyGlnAsnSerArgAsnHisValAspSerIleSerThrSerSerGluSerGln 480
 QY 3153 GCAATTAAGATCTGGAAGCAGCTGGCGTTGATTGGACAGAGTCACAAAAGGAAGCGAT 3212
 Db 481 AlaIleIySileLeuGluAlaIaIaGlyValAspLeuGlyGlnValThrIySgIySerAsp 500
 QY 3213 CCTGCGCTGACAACTGAAAAACAACATTGTATCACTGCAAGGAGTTAAGTTCCAGACGAA 3272
 Db 501 ProGlyLeuThrThrGluAsnAsnIleValSerLeuGlnGlyValIySValProAspGlu 520
 QY 3273 AACCTTACACCAACAAGCGGCAACATCGGGAAGAAACAGTTGCAAAAAATAAAAAATG 3332
 Db 521 AsnLeuThrProGlnGlnIuArgGlnHisArgGluGluGlnLeuAlaIySileIySblySmet 540
 QY 3333 AATCAATTTCTTTTCTCTGAAAAATGAGAAATTCAGTAGAGCTAATGTAACTCACAGATA 3392
 Db 541 AsnGlnPheLeuPheProGluAsnGluAsnSerValGlyAlaAsnValSerSerGlnIle 560
 QY 3393 ACAAAAATTCAGAGAGATTTAATGATGGGATGTCCGGTGGCGGAGCGGATCTATTATA 3452
 Db 561 ThrIySileProGlyAspLeuMetMetGlyMetSerGlyGlyGlyGlySerIleIle 580
 QY 3453 AATCGAGATGCGCAACTGCATATGCGAGTAAACGCCAAATCGAGCTCTTATCGCGG 3512

| | | | |
|----|------|--|------|
| Db | 581 | AsnProThrmMetArgGlnLeuHisMetProGlyAsnAlaIlysSerGluLeuLeuSerAla | 600 |
| QY | 3513 | ACAAGTTCAGAGCTTTCCGAAGATGTATGCATCCAGGGATGTTATATCAGATATGGGT | 3572 |
| Db | 601 | ThrSerSerGlyLeuSerGlyuAspValMetHisProGlyAspValIleSerAspMetGly | 620 |
| QY | 3573 | GCCGTAATAGATGTATATATATCAAAAAACCAGTGTGCAATGTGGATCTGGAGTAGGT | 3632 |
| Db | 621 | AlaValIleGlyCysAsnAsnAsnGlnLysThrSerValGlnCysGlySerGlyValGly | 640 |
| QY | 3633 | GTTGTCACTGGAAACAACCTGCACTGGAGTAAATGTCAATATGCAATTGCTCAAGCTCCGGC | 3692 |
| Db | 641 | ValValThrGlyThrThrAlaAlaGlyValAsnValAsnMetHisCysSerSerSerGly | 660 |
| QY | 3693 | GCCCCGAATGGCAATATGATGGGAAGCTCTACGAGATATGCTAGCCTCGTTGGCAACACA | 3752 |
| Db | 661 | AlaProAsnGlyAsnMetMetGlySerSerThrAspMetLeuAlaSerPheGlyAsnThr | 680 |
| QY | 3753 | AGCTGCAACGTCATCGGAACGGCCCCAGATATATGTTCTAAGAACTTTTAATCAAGATAGC | 3812 |
| Db | 681 | SerCysAsnValIleGlyThrAlaProAspMetSerLysGluValLeuAsnGlnAspSer | 700 |
| QY | 3813 | CGAACCATTTCACATCAAGGGGGAGTTGCTCAATGTAGTGTGGAAGATTCAACATCAA | 3872 |
| Db | 701 | ArgThrHisSerHisGlnGlyValAlaAlaGlnMetGluTrpSerLysIleGlnHisGln | 720 |
| QY | 3873 | TTTTTGAAGAACGCCCTCAAGGGGGGCAAGCCCCAGACAACTCACTGGAACGTGATACCA | 3932 |
| Db | 721 | PhePheGlnGluArgLeuLysGlyLysProArgGlnValThrGlyThrValValPro | 740 |
| QY | 3933 | CAACAGCAAAACCCCTTCTGGAATCTGTGTGAAACTCGTTAAACAACAGGTGCGACCCCTG | 3992 |
| Db | 741 | GlnGlnGlnThrProSerGlySerGlyGlyAsnSerLeuAsnAsnGlnValArgProLeu | 760 |
| QY | 3993 | CAAGTCCACCTCCTCCTTACCACCTCCATCCAGAGATCTGCTCAGTACCATAAGCACT | 4052 |
| Db | 761 | GlnGlyProProProProGlyHisSerIleGlnArgSerAlaSerValProIleAlaThr | 780 |
| QY | 4053 | CAATGCCCAATCCCTCGAGTCCAAACAATCTATCTCTCCGTCACCGCGAACAACGCA | 4112 |
| Db | 781 | GlnSerProAsnProSerSerProAsnAsnLeuSerLeuProSerProArgThrThrAla | 800 |
| QY | 4113 | GCAGTCATGGATTGCCGACAACCTCTCCTAGCATGGAAGGAACAGATCATTAATCTGA | 4172 |
| Db | 801 | AlaValMetGlyLeuProThrAsnSerProSerMetAspGlyThrGlySerLeuSerGly | 820 |
| QY | 4173 | TCTGTCCGCAAGCTAATACTTGCACGGTTCAGGCAAGGACAACAACAAGTGTCTACGA | 4232 |
| Db | 821 | SerValProGlnAlaAsnThrSerThrValGlnAlaGlyThrThrValLeuSerAla | 840 |
| QY | 4233 | AACAAGAACTGTTTTCAGGCAACACCCCATCGCCGTCAAATCAAAATCGTAGTAGAAT | 4292 |
| Db | 841 | AsnLysAsnCysPheGlnAlaAspThrProSerProSerAsnGlnAsnArgSerArgAsn | 860 |
| QY | 4293 | ACCGGATCGTCAAGCGTTCTTAACCATTAACCTTAAGCAGCAACCCAAGTACCCCTTATCT | 4352 |
| Db | 861 | ThrGlySerSerSerValLeuThrHisAsnLeuSerSerAsnProSerThrProLeuSer | 880 |
| QY | 4353 | CATCTATCCCCCAAGAAATTGAGTCTTTTCGTCAGTCCCTCTGCTGTATGTTATATTG | 4412 |
| Db | 881 | HisLeuSerProLysGluPheGluSerPheGlyGlnSerSer----- | 894 |
| QY | 4413 | TTTAATTTTTTAAAGACAATCAATATGAATTGCGTTAATAATAAGTTATATATTACA | 4472 |
| Db | 894 | ----- | 894 |
| QY | 4473 | TAACTCGGAATTTGATAGAAAAATCAGGAATAGAAAAAATAAATTATTTCCGACCG | 4532 |
| Db | 894 | ----- | 894 |
| QY | 4533 | CCCATCCATTCTTGAATCCAATTTCTGAGTAGATTGTTAGAGATAATCTACTATTAAAA | 4592 |

[illegible]

QY 5673 TACTCTGGAATGCCAGATTGATGGCGGACAGAGCCGAGGTAATGATTC 5732
 DB 1208 yThrProGlyMetProGlyLeuMetAlaGlyProGlyAlaGlyGlyMetLeuMetAsnSe 1228
 QY 5733 TTCCGAGAGCAACACAGAGATGACAAACATCCCTGGGGCAAGCATGGTATTAA 5792
 DB 1228 rSerGlyGlnGlnHisGlnAsnLysIleThrAsnAsnProGlyAlaSerAsnGlyIleAs 1248
 QY 5793 CTCTTTTCAAGATTGCAATCAATGTCTATTGTGACGAGAGGGTGATTAACCGGCCA 5852
 DB 1248 nPhePheGlnAsnCysAsnGlnMetSerIleValAspGlyGlyLeuProGlyHn 1268
 QY 5853 TGACGGATCAATGAATATTGGTCAACCATCTATGATAAGGGCATGCGTCCACATGCCAT 5912
 DB 1268 sAspGlySerMetAsnIleGlyGlnProSerMetIleArgGlyMetArgProHisAlaMe 1288
 QY 5913 GCGGCCAAATGTATGGGTGGCGGAGTGGCAACCGTTAAACAGGCAAAATTCAGTTTGACA 5972
 DB 1288 tArgProAsnValMetGlyAlaArgMetProProValAsnArgGlnIleGlnPheAlaG 1308
 QY 5973 GTCATCGGATGTAATTGACTGTGTGGGGATCCGTCATCATTTTTCATTAACGCTTCCTG 6032
 DB 1308 nSerSerAspGlyIleAspCysValGlyAspProSerSerPhePheThrAsnAlaSerCy 1328
 QY 6033 CAACAGCGCTGGACCAACATGTTTGATGACGACAAACAGGCGCAATCAGCTTAAGACACA 6092
 DB 1328 sAsnSerAlaGlyProHisMetPheGlySerAlaGlnGlnAlaAsnGlnProLysThrG 1348
 QY 6093 ACACATAAAGACATACCTAGTGAATGTGTCAAAACCAATCGGAGCTTGACAGTGGCACA 6152
 DB 1348 nHisIleLysAsnIleProSerGlyMetCysGlnAsnGlnSerGlyLeuAlaValAlaG 1368
 QY 6153 AGGCGAGATCCAACTGATGGGCAAGGACATGCGGAGTCACTTTTAATTGACCTAC 6212
 DB 1368 nGlyGlnIleGlnLeuHisGlyGlnGlyHisAlaGlnGlyGlnSerLeuIleGlyProTh 1388
 QY 6213 TAATATATATTAATGTCAACTGCCGGAAGTGTCACTACTAAACGGTGTCTTGGCAT 6272
 DB 1388 rAsnAsnAsnLeuMetSerThrAlaGlySerValSerAlaThrAsnGlyValSerGlyI 1408
 QY 6273 CAATTTGAGTCCCTCTTCTACGAGCTGAGTATGCCCCAGCAATATCATAGTTTCA 6332
 DB 1408 eAsnPheValGlyProSerSerThrAspLeuLysTyrAlaGlnGlnIleGlyHisSerPheG 1428
 QY 6333 GCAGCAGTTATATGCTTACCAACAGCAGAGTCAACAACAGCATATGACAGCAGCA 6392
 DB 1428 nGlnGlnLeuTyrAlaThrAsnThrArgSerGlnGlnGlnGlnHisMetHisGlnGlnH 1448
 QY 6393 CCAGAGCAACATGATTAACATGCGCGCGCAATTTATCCCAATCCACGTTCTTGTCAA 6452
 DB 1448 sGlnSerAsnMetIleThrMetProProAsnLeuSerProAsnProThrPhePheValAs 1468
 QY 6453 CAAA 6456
 DB 1468 nLys 1469
 RESULT 2
 Q67FY3
 ID Q67FY3 PRELIMINARY; PRT; 1530 AA.
 AC Q67FY3;
 DT 25-OCT-2004 (TrEMBLrel. 28, Created)
 DT 25-OCT-2004 (TrEMBLrel. 28, last sequence update)
 DT 25-OCT-2004 (TrEMBLrel. 28, last annotation update)
 DE Bcl9-2.
 OS Brachydanio rerio (Zebrafish) (Danio rerio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 OX NCBI_TaxID=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Brembeck F.H., Schwarz-Romond T., Bakkers J., Wilhelm S.,

RA Hammerschmidt M., Bircheimer W.;
 RT "Essential role of BCL9-2 in the switch between [beta]-catenin's
 RT adhesive and transcriptional functions.";
 RL Genes Dev. 18:0-0(2004).
 DR EMBL; AY296057; AAQ62695.1;
 SQ SEQUENCE 1530 AA; 159872 MW; C29FEC9433ED28C0 CRC64;
 Alignment Scores:
 Pred. No.: 9.93e-14 Length: 1530
 Score: 338.50 Matches: 308
 Percent Similarity: 30.82% Conservative: 171
 Best Local Similarity: 19.82% Mismatches: 554
 Query Match: 2.81% Indels: 522
 DB: 2 Gaps: 71
 US-10-664-859-1 (1-6909) x Q67FY3 (1-1530)
 QY 2565 CCAATACCCGGAATCGAGTTGAGCGGCGGAGAAATTTATGACTGCAACGCCAAT 2624
 DB 153 ProSerGlnGlnHisSerGlySerSerThrThrGlnHisValSerAspProGlyLysPro 172
 QY 2625 GGAATCTCTCGGAGTAGAGT-----AATGTTTGATTAATGCAACAGCAA 2672
 DB 173 GlyLeuGlySerGlyHisGlyProGlyIleArgThrAspLeuHisSerArgProProGln 192
 QY 2673 AATCACATATTCGTGTTTCAACTAGCTGGCCAAACAGGGCGGCAATCAGTTTAAAGC 2732
 DB 193 GlnValValTyrValPheThrThrXSerLeuAlaAsnSerAlaAlaGlyAlaValMetHis 212
 QY 2733 GGTCAATTTCAAACTATTATGCGTATGCACTGCACTAGCCTGTGTAACAAAGCTTCCTG 2792
 DB 213 GlnHisThrAspSerIleLeuLeuYrHis--GlnGlnAsnValProArgThrLysLeu 231
 QY 2793 GAAAGCTTTTATGAAAAACCTTTAAAGATTAAAGATTAAAGGCGGCAATTCCTGTC 2852
 DB 232 AspGln-----SerThrGlyValGlyLysValSerAsnLeuAlaGlnHisIleSerSer 249
 QY 2853 GGTATGCCATGATAGGCATGCGGCGGAGTTGAGTAACTCTCTGTAATCTGTAGCCAA 2912
 DB 250 SerHis-----SerProProIleGlyThrProLys 259
 QY 2913 ATAACACAAACAGCAGCCACATACAAAGACCGTAGGCTATTGAAACCCCAATTCATCA 2972
 DB 260 SerGlnSerGlyThrProArgProAlaSerValGlyGlyVal----- 273
 QY 2973 CATGAAACAGCAAAAGCTAGTACTGTAAAGCGGCTAGCAACTCTTTTGTGCAACAGTCT 3032
 DB 274 -----GlnHisLeuProGlyThrSerThrProSerSerThrGlnHisProAspSer 290
 QY 3033 GATCCTATGGGCAAGCAACTGAATGTGTGCTGGGAGCGGAGCTCTCAAAACAGCAGT 3092
 DB 291 GluProAlaGlnThrHis-----ArgGlyGlyGlyThr----- 301
 QY 3093 AGTCTGCAAAAACCTCAGCAATCATGTAGACAGTATCAGTACA---TCCAGCGAGTCA 3149
 DB 302 ---SerSerAsnAsnGlnLysArgSerAlaValHisThrLeuGlyLeuGlyAsnSerGlyPro 320
 QY 3150 CAGGCAATTAAGATTAAGAGCAGCTGCGCTGATTTGGGACAGGTTACAAAGGAAAGC 3209
 DB 321 GlnSerValGlyValSerGlyThrGlnGlyValAspArgProGlyAlaIleProHisHis 340
 QY 3210 GATCCTGGCTGACAACTGAAACAAACATGTATCACTGCAAGGAGTTAAGGTTCCAGAC 3269
 DB 341 GlyAlaGlyValSerProSerThrSerProSerValLeuSerAlaLeuArgGlnSerGly 360
 QY 3270 -----GAAAACCTTACACCAACAAGCGGCAACAT 3299
 DB 361 LeuGlyGlnArgValGlyProGlyAsnThrAspGlyLeuSerLysGlnGlnLeuGlnHis 380
 QY 3300 CGGAAGAACAGTTGGCAAAATAAATAAATAATGAATCAATTTCTTTCTGAAAAATGAG 3359
 DB 381 ArgGluArgSerLeuGlnThrLeuArgAspIleGluArgLeuLeuArgSer----- 398

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: November 2, 2005, 23:46:44 ; Search time 206 Seconds
(without alignments)
6453.996 Million cell updates/sec

Title: US-10-664-859-1

Perfect score: 12037
Sequence: 1 acgagtccttccttattat.....aaaaaaaaaaaaaaaaaaaa 6909

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283416 segs, 96216763 residues

Total number of hits satisfying chosen parameters: 566832

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10664859/runat_02112005_171808_3665/app_query.fasta_1.7047
-DB=PIR -QFMT=fastan -SUFFIX=rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10664859_@CGN_1_1_338_@runat_02112005_171808_3665 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR_79:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|----------|----------------------|
| 1 | 301 | 2.5 | 2232 | 2 T34434 | hypothetical prote |
| 2 | 270.5 | 2.2 | 1596 | 2 A33106 | neurogenic locus m |
| 3 | 258.5 | 2.1 | 1655 | 2 T13998 | gene mastermind pr |
| 4 | 249.5 | 2.1 | 2038 | 2 A43742 | female sterility hom |
| 5 | 238.5 | 2.0 | 2715 | 2 T13049 | eyelid - fruit fly |
| 6 | 233 | 1.9 | 3190 | 2 T13828 | CREB-binding prote |
| 7 | 231.5 | 1.9 | 1844 | 2 D71612 | hypothetical prote |
| 8 | 228 | 1.9 | 1424 | 2 T03851 | thyroid hormone re |
| 9 | 222.5 | 1.8 | 1063 | 2 D86731 | hypothetical prote |
| 10 | 222.5 | 1.8 | 1459 | 2 T32271 | hypothetical prote |
| 11 | 219 | 1.8 | 1457 | 2 T14577 | protein kinase Yak |
| 12 | 216.5 | 1.8 | 1061 | 2 A57620 | steroid receptor c |
| 13 | 215 | 1.8 | 786 | 2 T16509 | hypothetical prote |
| 14 | 214.5 | 1.8 | 1142 | 2 T00022 | B120 protein - hum |

| | | | | | |
|----|-------|-----|-------|----------|--------------------|
| 15 | 213 | 1.8 | 925 | 2 T19361 | hypothetical prote |
| 16 | 207 | 1.7 | 1441 | 2 B86807 | hypothetical prote |
| 17 | 206 | 1.7 | 1088 | 2 H96747 | unknown protein T1 |
| 18 | 205.5 | 1.7 | 959 | 2 B44402 | nuclear pore compl |
| 19 | 205 | 1.7 | 1237 | 2 D71850 | probable outer mem |
| 20 | 204 | 1.7 | 1584 | 2 T18276 | protein-tyrosine k |
| 21 | 203.5 | 1.7 | 13288 | 2 T03099 | mucin, submaxillar |
| 22 | 201.5 | 1.7 | 1032 | 2 T34433 | hypothetical prote |
| 23 | 201.5 | 1.7 | 1436 | 2 D71618 | hypothetical prote |
| 24 | 201 | 1.7 | 1778 | 2 T50074 | probable nucleopor |
| 25 | 200.5 | 1.7 | 907 | 2 A45560 | sporozoite surface |
| 26 | 198 | 1.6 | 1230 | 2 B64664 | outer membrane pro |
| 27 | 197 | 1.6 | 1532 | 2 A61262 | collagen alpha 1(X |
| 28 | 195.5 | 1.6 | 3968 | 2 A44265 | trithorax homolog |
| 29 | 194.5 | 1.6 | 1436 | 2 S57238 | forked protein 5.4 |
| 30 | 194.5 | 1.6 | 1449 | 2 S57237 | forked protein 5.6 |
| 31 | 193 | 1.6 | 1113 | 2 S28925 | nuclear pore compl |
| 32 | 193 | 1.6 | 2468 | 2 A83412 | hypothetical prote |
| 33 | 193 | 1.6 | 2893 | 2 A64556 | toxin-like outer m |
| 34 | 192 | 1.6 | 2703 | 1 A24420 | notch protein - fr |
| 35 | 191.5 | 1.6 | 2902 | 2 C71953 | hypothetical prote |
| 36 | 190 | 1.6 | 1307 | 2 T25563 | sericinB - silkw |
| 37 | 189 | 1.6 | 1217 | 2 S52714 | sensory transducti |
| 38 | 188.5 | 1.6 | 2150 | 2 S71629 | hypothetical prote |
| 39 | 188 | 1.6 | 1072 | 2 A86827 | collar protein iso |
| 40 | 188 | 1.6 | 1296 | 2 T13936 | hypothetical prote |
| 41 | 188 | 1.6 | 2027 | 2 S60123 | protein R10E1.1 l |
| 42 | 188 | 1.6 | 2056 | 2 G88564 | probable adhesin h |
| 43 | 186 | 1.5 | 1910 | 2 AF0394 | 1-phosphatidylinos |
| 44 | 185.5 | 1.5 | 1858 | 2 T18273 | hypothetical prote |
| 45 | 185.5 | 1.5 | 3507 | 2 T34513 | |

ALIGNMENTS

RESULT 1

hypothetical protein K06A9.1a - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004
C/Accession: T34434
R/Geisel, C.; Gattung, S.
submitted to the EMBL Data Library, December 1996
A/Description: The sequence of C. elegans cosmid K06A9.
A/Reference number: Z21525
A/Accession: T34434
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-2232 <GET>
A/Cross-references: UNIPROT:Q8IFX6; EMBL:U80846; PIDN:AAC70890.1; GSPDB:GN00028; CESP:K
A/Experimental source: strain Bristol N2; clone K06A9
C/Genetics:
A/Gene: CESP:K06A9.1a
A/Map position: X
A/Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 1088/1; 1367/1; 2039/1; 2049/1; 2075/

Alignment Scores:

| | | | |
|------------------------|----------|---------------|------|
| Pred. No.: | 6,48e-12 | Length: | 2232 |
| Score: | 301.00 | Matches: | 297 |
| Percent Similarity: | 33.17% | Conservative: | 187 |
| Best Local Similarity: | 20.36% | Mismatches: | 599 |
| Query Match: | 2.50% | Indels: | 378 |
| DB: | | Gaps: | 62 |

| | |
|---|--|
| US-10-664-859-1 (1-6909) x T34434 (1-2232) | |
| OY 2355 GTGTATTTTAATAATATCTTATTTATTCATTTTACATAGTTAAATTTGAAGATTTC A 2414 | |
| DB 174 ValGlnThrLysLysTyrValIleValSerLeu-----AlaMetAsp 187 | |
| OY 2415 AACGACAGTACGACGAGAAAAAAGATCGTCTTGACAATGAATAATGACGAATGAGC 2474 | |
| DB 188 GlnGlyAsnMetThrGlnLysTyrGlyAspLeuAsnThrHisAsnIleGlyAsnAsnAla 207 | |

QY 2475 ATGAAGGCTGCAATCAGTTGAATCCCGATTTTATCAATGAATCTTTAATAATCCTGCA 2534
Db 208 ThrAsp1leThrAsnAsp1leAsnValleuAsn1leThrAla1aThrSerProser 227
QY 2535 ATTCGAGCATATTAGTAGCGGAGTAGA-----CCAAATACCCCGAATCCGA 2582
Db 228 ThrAlaVal1ThrSerProserSerleuGlyThrSerSerSerProleuProser----- 245
QY 2583 GTTGAGCGGGACGGGAATTATTGACTGCCAACGCCAATGGAATCTCCTCGGTAGC 2642
Db 246 ---Ser1leSerThrSerAla1euPro1leAlaSerSerSerAlaSerSerProser 264
QY 2643 AGTAATTGTTGGATTACATGCAACAGCAAAATCACATATTGCTGTTTCACTCAGCTG 2702
Db 265 Ala-----AlaSerSerThrThrProValValleuSerSerThr1le 279
QY 2703 GCCAACAAAGGGCCGAATCAGTTTAAAGCGGTCAATTTCAACTATTATGCGTATCAC 2762
Db 280 GlnSerSer-----SerGlyThrPheProSerSerValAla----- 291
QY 2763 TGCACTCAGCCTGCTACAATAAAGCTTCGTGAAGAAGCTTTTATGAAAAACCTTAAAG 2822
Db 292 ---SerSerProserThrValGly-----Ser 299
QY 2823 ATTAACAAGTTACAGCGGCACACATTCGTCGATGCGATGATGAGCATGGGCAGGTT 2882
Db 300 ThrSerGlyAlaAlaSerSerSerSerGlyAlaThrValSerThr1leAlaGlySerThr 319
QY 2883 GGACTAACTCCTCTTAATCCTGTAGCCAAATAACACACACAGCAGCCACATACAAAGACC 2942
Db 320 GlySerThr-----1leThrProValProGlySerSerSerThr 332
QY 2943 GTAGGCTTATTGAAACCCCAATTCAATCACATGAAAAACAGCAACGTAAGTACTGTAGC 3002
Db 333 1leGlySerSerThrProSerAlaSerSerSerSerGlyThrMetSerThr1leSer 352
QY 3003 GCGCCTAGCACTCTTTGTGCAC-----CAGTCT 3032
Db 353 GlySerThrGlySerThrValThrValAlProGlySerSerSerThrPheAlaSerSer 372
QY 3033 GATCCTATGGGCAACGAACT-----GAATTGATGTCTGGAGCGGATCCTCA 3083
Db 373 ThrPro1leAlaSerSerSerSerProGlySerThrValThrValAlaProGlySerSer 392
QY 3084 AACACC-----AGTAGCTCTGCACAAAACTCACGAATCATGTAGACAGTATC 3131
Db 393 SerThrThrGlySerSerThrProSerAlaSerSerSerSerGlyThrMetSerThr 412
QY 3132 AGTACATCCACGAGTCAACAGCAATAAAGTACTGGAAGCAGCTGGCGTTGATTGGGA 3191
Db 413 AsnSerGlySerThrGlySerThrValThrValAlaProValSerSerSerThrPheGly 432
QY 3192 CAGGTACAC-----AAAGGAGCGATCCTGCGCTGACAACTGAAAAACAAC 3236
Db 433 SerSerThrPro1leAlaSerSerSerSerSerGlySerThrValThrValSerGly 452
QY 3237 ATTGTATCACTGCAAGAGTTAAGGTTCCAGACGAAAACTTACACCAACAGCGGCAA 3296
Db 453 SerSerSerThrThrGlySerSerThrProSerAlaSer-----SerSerSerAla 469
QY 3297 CATCGGGAAGACAGTTGGCAAAAATAAAAAATGAATCAATTTCTTTTCCCTGAAAT 3356
Db 470 GlyThrAlaSerThr1leSerGlySerThrGlySerThrAlaThr1leValProGlySer 489
QY 3357 GAGAATTCAAGTAGAGCTAATGTAAGCTCACAGATTAACAAAAATTCAGAGAGATTATG 3416
Db 490 SerSerSerValGlySerSerThrGlnSerAlaSerProSerSerProGly---ThrMet 508
QY 3417 ATGGGAGATCGGGTGGCGGAGCGGATCTATTATAATCCGACGATGCGACAACCTGCAT 3476
Db 509 SerThrValSerGlyProThrGlySerThrVal-----ThrVal 521

QY 3477 ATGCCAGTAACGCCAAATCGGAGCTCTTATCGCGCAGCAAGTTCAAGACTTTCGGAAGAT 3536
Db 522 ValProGlySerSerThrSerProAlaProSerSerSerPro----- 535
QY 3537 GTAATGATCCAGGGAGTGTATATCAGATATGCGTCCGTAATAGATGTAATAAT 3596
Db 536 ---AsnProSerSerSerProAlaSerThrGlySerThr1leThr1leSerGlySer 553
QY 3597 CAAAAAACAGTGTCAATGTGATGTGAGTAGGTGTTGTCACTGGAACAACCTGACAGCT 3656
Db 554 SerSer1le1leValSerThrValSerGly---SerThrValSerGlySerThr----- 570
QY 3657 GGAGTAAATGTCAATATGATGCTCAAGCTCCGGCGCCCG---AATGCAATATGATG 3713
Db 571 GlyThrSerGlnSerThr1leuAlaSerSerThrAlaThrProGlySerSerSerThrVal 590
QY 3714 GGAAGCTCTACGGATATGCTAGCCTGTTGGCAACACAGCTGCAACGTCATCGGAACG 3773
Db 591 ProSerSerSerSerProGlnProSerSerGlnSerProAlaProAsnThrGlySerThr 610
QY 3774 GCCCCAGTATGCTTAAGAA-----GTTTAAATCAAGATAGCCGAACC 3818
Db 611 ThrProSerGlnThrSerSerGlnSerProSerProSerMetAsnProSerSerSerThr 630
QY 3819 CATTCACATCAAGGGGAGTGTCTCAATGAGTGTGGAAGATTCAACATCAATTTTTC 3878
Db 631 ProThr-----GlySerSerGlnSerThr1leThr 640
QY 3879 GAAGAAGCCTCAAGGGGGCAAGCCCAAGCAAGTCACTGAACT----- 3923
Db 641 ProGlyGlySerThrAlaSerSerProThrGlySerThrGlySerThrPheSerValAla 660
QY 3924 -----GTAGTACCACACACAGCAAAACCCCTTCTGATCT-----GGTGAAACTCGTTA 3971
Db 661 ThrGluValThrSerGlnSerThrValProSerGlySerSerleuGlyThrGlnSerThr 680
QY 3972 AACACCAAGGTGCGACCCCTGCAAGGTCCACCTCCTTACCACCTCATCCAGAGATCT 4031
Db 681 AsnSerSer-----ProSerProSerSerleuSerProSerThr 693
QY 4032 GCGTCAATCAATAGCCCACTCAATCGCCCAATCCCTGAGTCCAAACAATCTATCTCTC 4091
Db 694 SerGlyMetSerThrLeuThrSerGluProSerProSerSerThrGlnSerSerGlyAla 713
QY 4092 CCGTACCGCGGACAACCGCAGCATGAGATTGCCGACCACTCTCCTAGCATGAT 4151
Db 714 GlnSerThrLeuThrThr-----ProSerProAsnProSer---Gln 726
QY 4152 GGAACAGATCATTATCTGATCTGTCCGCAAGCTAATACTTGACGCGTTCAGCAGGC 4211
Db 727 SerThrSerSerleuGluSerSerThrSerGlyAlaThrThrSerSerGlySerAlaGly 746
QY 4212 ACAACACAGTGTCTCAGCAACAAGAACTGTTTCAGGCAGACACCCCATCGCGCTCA 4271
Db 747 ThrThrMetThr-----SerProSer 753
QY 4272 AATCAAAATCGTAGAATAACCGGATCGTCAAGCGTCTTACGCAATACTTAAGCAGC 4331
Db 754 -----GlnSerSerSerValGlySerSerGlnGlySerThrSerProAlaAlaSer 770
QY 4332 AACCCAGT-----ACCCCTTATCTCATCTATCC 4361
Db 771 ThrThrSerGlyGluMetThrSerGlnGlySerThrGlnThrProGlySerSerValSer 790
QY 4362 CCAAGGAATTGAGTCTTTCGTCAGTCTCTGCTGTGATGTTATATTGTTAATTTT 4421
Db 791 -----ThrSerAlaAla1leu----- 796
QY 4422 TTTAAAGACAATCAATATGAATGCGTTAATAATAAGTTATATATACATAACTCGGA 4481
Db 797 -----ThrSerThrGlnGlnSerValSerThrAsnSer-----ProGly 809
QY 4482 AATTGATAGAAAAAATCAGGAATAGAAAAAATAATTATTTCCGAGCCGCATCAT 4541